Universidad Autónoma de Madrid

Departamento de Bioquímica

Doctoral Thesis

Search for new susceptibility genes in hereditary BRCAX families with an apparent recessive pattern of inheritance using Whole Exome Sequencing

Alejandra Tavera Tapia

Madrid, 2017

Universidad Autónoma de Madrid
Facultad de Medicina
Departamento de bioquímica

Search for new susceptibility genes in hereditary BRCAX families with an apparent recessive pattern of inheritance using Whole Exome Sequencing

---

Doctoral Thesis

MSc in Cell Biology

Alejandra Tavera Tapia

Thesis Directors
Dr. Ana Osorio Cabrero
Dr. Javier Benítez Ortiz

Human Genetics Group
Human Cancer Genetics Programme
Spanish National Cancer Research Centre

This thesis, submitted for the degree of Doctor of Philosophy at the Autonomous University of Madrid, has been elaborated in the Human Cancer Genetics laboratory at the Spanish National Cancer Research Center (CNIO), under the supervision of **Dr. Ana Osorio Cabrero** and **Dr. Javier Benítez Ortíz**.

This research was supported by the following grants and fellowships:

- La Caixa/CNIO International PhD Fellowship, 2013-2017: Alejandra Tavera Tapia
- Boehringer Ingelheim Fonds Travel Fellowship, 2016: Alejandra Tavera Tapia
- The Spanish Ministry of Economy and Competitiveness (MINECO) SAF2014-57680-R. The Spanish Health Ministry PI12/00070, FEDER funds and H2020 (BRIDGES project, number 634935), and FIS 15/00059
- Project "300 exomes to elucidate rare diseases" by National Center for Genomic Analysis (CNAG)
- CIBERER

Summary

Only two genes involved in high-risk breast/ovarian cancer hereditary syndrome (HBOC) have been identified, *BRCA1* and *BRCA2*, that do not explain more than 20% of all the HBOC cases. In the last years, Whole Exome Sequencing (WES) has been used to find new high susceptibility genes. Although initial attempts were not as conclusive as expected; recently, novel susceptibility genes for HBOC have been reported using WES. Thus, the main objective of this study is the quest for new high susceptibility genes in BRCAX hereditary breast cancer families with an apparent recessive pattern of inheritance, which has remained largely unexplored. After reviewing about 2000 stories of familial breast cancer, we selected four families showing an apparent pattern of monogenic recessive inheritance with the following characteristics: presence of two or more affected siblings with breast cancer at young age, absence of familial antecedents of the disease, availability of the samples and no mutations in *BRCA1* and *BRCA2* genes.

In our first approach, studying the recessive model of inheritance, we found 14 candidate variants in two families where the exomes from the parents and the affected siblings was accessible (family 1 and 2). After that, regardless of the inheritance model, we explored the genes related with DNA repairing and maintenance systems, where we found a novel mutation in *ATM* (c.5441delT; p.Leu1814Trpfs*14), an already recognized gene for hereditary breast cancer, as a causal variant in a BRCAX family. We found other family with variants affecting *RECQL5* gene (1709C>T, p. T570I and c.2874C>G, p. S958R), an excellent candidate to be associated with the disease. As we had found at least one family where the dominant model explained the cause of the disease, we completed the panorama by studying the rest of the genes under this model in the three families. We ended up with a total of 43 candidate variants from which we finally selected 25 as putative susceptibility alleles for hereditary breast cancer after performing a case-control association study in 1,500 cases of BRCAX families and 500 controls in Spanish population.

We performed full coding sequencing and exon-boundaries analysis of *RECQL5* and *ATM*, in order to establish the frequency and spectrum of mutations of both genes in Spanish population. Regarding *RECQL5*, Truseq platform from Illumina was used in a cohort of 700 BRCAX BC-only cases and 754 controls. From this study, we propose *RECQL5* as a novel BC susceptibility gene, by the identification of clearly deleterious mutations in the BRCAX cohort studied. Other potentially deleterious mutations were identified and *in silico* studies predict them to affect preferentially helicase domain of this enzyme. Concerning *ATM*, Next Generation Sequencing panels were used for *ATM* mutational screening in a cohort of 392 HBOC Spanish BRCAX families and 350 controls. 1.78% prevalence of mutations in the *ATM* gene was associated to HBOC and 1.94% in breast cancer-only BRCAX families in Spanish population. In the fourth family, which is a family with various cases of male breast cancer (MBC), we explored the recessive model and the X-linked inheritance model. Seven candidate variants were identified as putative susceptibility alleles for MBC, which were studied in an initial cohort of 50 Spanish hereditary MBC cases and in an enlarged cohort of 1200 MBC and 500 controls from UK. In the Spanish cohort, we found another MBC case that was homozygous for c.1208G>A p.R403Q affecting *TXNDC5* gene which is involved in the Androgen Receptor pathway and therefore we propose as a novel putative susceptibility allele for MBC, that requires further study. In conclusion, these set of studies using WES for exploring HBOC missing hereditability has guided to the exploration of different models of inheritance and has been useful to find new BC susceptibility candidate genes, as well as variants not previously reported in well-established BC susceptibility genes, which contributes to the understating of the still uncovered BRCAX landscape.

Resumen

Sólo dos genes que confieren alta susceptibilidad al síndrome de cáncer de mama y ovario hereditario (CMOH) han sido identificados, *BRCA1* y *BRCA2*, que no explican más del 20% de todos los casos de CMOH. La secuenciación completa de exoma ha sido empleada para encontrar nuevos genes que confieran alta susceptibilidad a HBOC. Así, el objetivo principal de este estudio es la búsqueda de nuevos genes de alta susceptibilidad en familias de cáncer de mama hereditario BRCAX con un patrón aparente de herencia recesiva mediante el uso de secuenciación de exoma completo. Después de revisar 2000 historias clínicas de cáncer de mama familiar, hemos seleccionado cuatro familias que muestran un patrón aparente de herencia monogénica recesiva: presencia de dos o más hermanas afectadas con cáncer de mama en edad joven, ausencia de antecedentes familiares de la enfermedad y ninguna mutación en los genes *BRCA1* y *BRCA2* (BRCAX).

Estudiando el modelo de herencia recesivo, encontramos 14 variantes candidatas en dos familias donde los exomas de los padres y las hijas afectas con cáncer de mama estaban disponibles (familia 1 y 2). Después, independientemente del modelo de herencia, exploramos genes relacionados con sistemas de reparación y mantenimiento del DNA, donde encontramos una mutación no reportada previamente en *ATM* (c.5441delT; p.Leu1814Trpfs*14), un gen previamente reconocido como un gen de susceptibilidad a cáncer de mama, como la variante causal en una familia BRCAX. También, encontramos otra familia con dos variantes en el gen *RECQL5* (1709C>T, p. T570I y c.2874C>G, p.S958R), que consideramos un excelente candidato posiblemente relacionado con la enfermedad. Dado que encontramos al menos una familia donde el modelo dominante explicaba la causa de la enfermedad, completamos el panorama estudiando el resto de genes bajo este modelo de herencia en las tres familias. De las 43 variantes candidatas iniciales, identificamos 25 como alelos putativos de susceptibilidad a cáncer de mama hereditario, seleccionadas a través de un estudio de asociación de casos y controles efectuado en 1500 casos de familias BRCAX y 500 controles de población española.

Realizamos la secuenciación completa de regiones codificantes en *RECQL5* y *ATM*, a fin de establecer la frecuencia y espectro de las mutaciones de ambos genes en población española. En cuanto a *RECQL5*, el gen fue analizado utilizando la plataforma Truseq de Illumina en una cohorte de 700 casos de mama de familias BRCAX y 754 controles. De este estudio, proponemos *RECQL5* como un nuevo gen que confiere susceptibilidad a cáncer de mama por la identificación de mutaciones claramente deletéreas en la cohorte de casos BRCAX. Otras mutaciones potencialmente deletéreas fueron identificadas y los estudios *in silico* predicen que afectan preferencialmente al dominio helicasa de esta enzima. Con respecto a *ATM*, se usó un panel de secuenciación de nueva generación para su análisis mutacional en una cohorte de 392 CMOH de familias BRCAX y 350 controles. Se encontró una prevalencia de 1.78% de mutaciones el gen *ATM* asociada con CMOH y 1.94% asociada a cáncer de mama en familias BRCAX en población española. En la cuarta familia, caracterizada por la presencia de varios casos de cáncer de mama en varón (CMV), exploramos el modelo recesivo y de herencia ligado al X. Siete variantes candidatas fueron identificadas como alelos putativos de susceptibilidad a CMV, las cuales fueron estudiados en una cohorte inicial de 50 casos de CMV hereditarios de población española y en una cohorte de 1200 CMV y 500 controles de población inglesa. En población española, encontramos otro caso de CMV que era homocigoto para c.1208G>A p.R403Q, afectando el gen *TXNDC5*, relacionado con la ruta del receptor de andrógenos, el cual proponemos como un gen candidato de susceptibilidad al CMV hereditario que requiere más estudios moleculares. En conclusión, este conjunto de estudios usando la secuenciación de exoma completo ha sido útil para encontrar nuevos genes candidatos que confieren susceptibilidad a cáncer de mama, así como variantes no reportadas previamente en genes de susceptibilidad a cáncer de mama ya establecidos como tales, lo cual contribuye a entender una parte del panorama aún irresuelto de los casos BRCAX.

Table of contents

Table of contents

1. Abbreviations 23

2. Introduction 27

2.1. Breast cancer

2.1.1 Mammary gland anatomy and breast cancer 29

2.1.2 Epidemiology 34

2.1.3 Risk factors 34

2.2. Hereditary breast cancer 37

2.2.1 High susceptibility genes 38

2.2.2 Moderate susceptibility genes 39

2.2.3. Low susceptibility genes 40

2.3. Next generation sequencing 41

2.3.1. Brief concept into NGS 41

2.3.2. Whole Exome Sequencing in hereditary breast cancer 43

2.3.3. Limitations of WES 47

2.3.4. Clinical Relevance 48

3. General Objective 51

4. Patients and Methods 55

4.1 Patients and Families of study 57

4.2 Methods 61

4.2.1 Whole Exome Sequencing 61

4.2.2 Bioinformatic analysis and variant filtration 61

4.2.3 Segregation analysis and Sanger validation 63

4.2.4 Case-Control Association Study 63

4.2.5 Splicing studies 63

4.2.6 Loss of heterozigosity (LOH) analysis 64

4.2.7 Full coding sequencing of *RECQL5* 64

4.2.8. *In silico* inference of missense variants effect in RECQL5 domains 65

4.2.9 Mutational analysis of *ATM* gene 66

4.2.10 Immunohistochemistry 67
4.11 Genotyping of 50 Spanish MBC 67
4.12 (KASP) Genotyping technology 67

5. Results 69
5. 1. Exploration of the recessive model of inheritance 71
5. 2. Exploration of genes related with DNA repairing systems 73
5. 3. Exploration of the dominant model of inheritance in BRCAX families 78
5. 4. Putative susceptibility genes associated with HBOC (BRCAx) families identified through WES in Spanish population 80
5.5 Massive sequencing of *RECQL5* gene in Spanish population 83
5.6 Almost 2% of Spanish BC families are associated to germline pathogenic mutations in *ATM* gene 95
5. 7. Search for novel susceptibility genes for hereditary male breast cancer 97

6. Discussion 101
6. 1. Exploration of different models of inheritance in BRCAX families 103
6. 2. Main findings in family BRCAX 1: Best candidates putatively associated with increased susceptibility to hereditary BC 105
6. 3. Main findings in family BRCAX 2:
6.3.1. *RECQL5*, another DNA helicase potentially involved in increased BC susceptibility 107
6.3.2. Other findings in family 2 exploring dominant and recessive model of inheritance 112
6. 4. Main findings in family BRCAX 3: Almost 2% of Spanish BC families are associated to germline pathogenic mutations in the *ATM* gene 114
6. 5. Main findings in family BRCAX 4: Novel susceptibility alleles potentially implicated with hereditary male breast cancer 115

7. Conclusions 119

8. Supplementary Data and Tables 127

9. References 143

10. Publications and poster presentations 161

1. Abbreviations

ATM	Ataxia Telangiectasia Mutated
BC	Breast Cancer
BRCA1	Breast Cancer Susceptibility Gene 1
BRCA2	Breast Cancer Susceptibility Gene 2
BRCAX	Breast Cancer Susceptibility Gene X
cDNA	Complementary DNA
CHEK2	Checkpoint kinase 2
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleic triphosphate
DSB	DNA double-strand break
FDA	Food and Drugs Administration from the United States
FFPE	Formalin Fixed Paraffin Embedded
GMAF	Global Minor Allele Frequency
GWAS	Genome Wide Association Study
FA	Fanconi Anemia
HBOC	Hereditary Breast and Ovarian Cancer
HR	Homologous Recombination
IHC	Immunohistochemistry
KASP	KBioscience Competitive Allele-Specific PCR
KEGG	Kyoto Encicopledia of Genes and Genomes
LOH	Loss of heterozigosity
MBC	Male Breast Cancer
mRNA	Messenger Ribonucleic acid
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NMD	Nonsense-mediated mRNA decay
PALB2	Parter and localizer of BRCA2
PARP	Poly ADP (Adenosine Diphosphate) Ribose Polymerase
PCR	Polymerase Chein Reaction
PTC	Premature termination codon
RAD51	Recombination protein A 51
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single Nucleotide Polymorphism
SSB	DNA single-strand break
STK11	Serie/threonine kinase 11
TEB	Terminal end bud
TP53	Tumor Protein 53
UTR	Untranslated region
VUS	Variants of unknown significance
WES	Whole Exome Sequencing
WT	Wild type

2. Introduction

2.1. Breast cancer

2.1.1 Mammary gland anatomy and breast cancer

Macroscopically, the mammary gland in adult women has between 15-20 lactiferous ducts, which begin at the nipple, branch into smaller ducts (segmentary and subsegmentary ducts) and culminate in the terminal duct lobular unit (lobule), which is composed of a terminal duct and many small ductules (or acini), which are the responsible of the production of maternal milk upon lactation period for feeding a newborn (Netter, 2007). This structure also possesses intralobulillar stroma (which separates the acini from each other, has loose connective tissue, lymphocytic infiltrate, capillary red blood and lymphatic vessels). The interlobulillar stroma has dense connective tissue rich in collagen and elastic fibers whereas the interlobular stroma has dense connective tissue surrounding the ramifications of the galactophoretic duct. There are also Cooper ligaments that are connective tissue strands that extend from the deep fascia to attach to the underlying skin and adipose tissue which determines the shape and size of the breast (Townsend et al., 2013) [Figure 1A].

Histologically, the main unit of the lobuli is named terminal end bud (TEB), which is a bulbous structure that has two main compartments, an outer one that is constituted by cap cells which differentiate into myoepithelial cells to allow the elongation of the duct; also they are theorized to be a reservoir of regenerative mammary stem cells. The inner part of the TEB has a multicellular layer of 4-6 cells, known as the body cells, which also harbor luminal and alveolar progenitors that upon differentiation give rise to the more mature luminal (predominate toward the neck of the TEB and the joint with the ducts) and alveolar cells (simple cubic tissue that are secretory cells, responsible for production of the milk). TEB structure is delimited by a basal lamina constituted of epithelial cells. There is a plethora of cellular types surrounding the laminar basement, among which we can find adipocytes, fibroblasts, macrophages, eosinophils, neutrophils and endothelial cells. Predominating cellular types within the ducts are the ones belonging to the basal lamina, myoephitelial cells (muscle contracting cells that are important for secretion of the content of the TEB toward the ductules) and luminal cells which form the body of the ducts (Paine et al., 2017; Ross et al., 2013) [Figure 1B].

The mammary glands have mesodermal origin and are formed since the fifth week of embryonic development. During childhood, the mammary glands do not develop and remain in relative quiescence until the onset of puberty. At that time, the tissue underneath the areola as well as the body of the gland increases in response to hormones such as progesterone, prolactin, corticoids and the growth hormone. The glandular ducts are scarcely branched (Netter, 2007).

Due to the influence of ovarian hormones, the mammary gland is modified cyclically, all of which increases its vascularization. After puberty, the tissue develops and remains in this form, until the stage of pregnancy, where other changes in the mammary gland are appreciated (Macias et al., 2012). Given the large amount of estrogen and progestogen during this stage, the mammary gland develops and branches considerably. Other modifications are appreciated such as: myoepithelial hypertrophy, hyperpigmentation and hypervascularization, as well as ductal and lobular development, at the ends of the ducts alveoli are formed where the milk accumulates. After lactation, the hormonal influence decreases and produces a regression of the breast by apoptosis, cellular degeneration and decrease in the number of alveoli and number of breast ducts (Lippert, 2005).

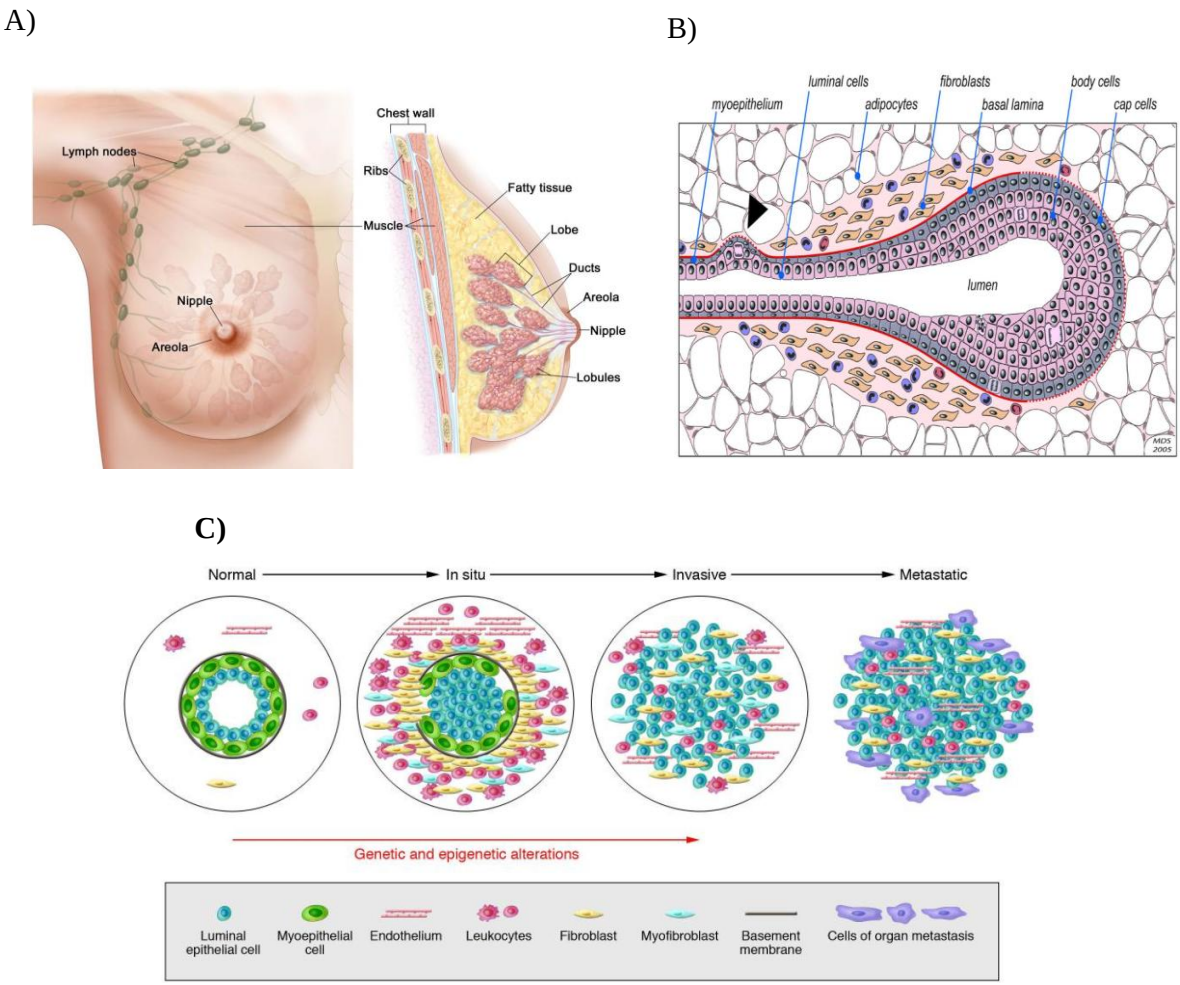


Figure 1. (A) Normal structure of adult mammary gland. (B) Histology of the mammary gland, the morphology of a duct and its terminal lobule together with the main cellular types that conform this structure are presented in the diagram. (C) Hypothetical model of breast cancer progression (Adapted from National Cancer Institute, 2016; Sternlich, 2006; Polyak, 2007).

Breast cancer (BC) is a highly complex disease, which is composed of distinct subtypes associated with different clinical outcome. Multiple factors have been associated to its development but the causes and mechanisms related have not been completely understood and several research lines have been established to improve current diagnosis, prognosis, treatment and prevention guidelines (Polyak, 2007).

There are different breast cancer types, which receive their name based in their location as well as grade of progression. After premalignant lesions occur and upon tumoral transformation, a carcinoma *in situ* develops, where cells have not surpassed the basal layer and remain inside of the ducts (so called ductal carcinoma) or the lobuli (lobular carcinoma). Then, in a more advanced stage, the carcinoma can acquire infiltrating properties by disrupting the basal layer and invading the lumen of the duct (so called infiltrating ductal carcinoma) or the lobuli (infiltrating lobular carcinoma). The infiltrating ductal carcinoma is the most common type and the infiltrating lobular carcinoma arises in 10-15% of all BC cases. The infiltrating carcinomas in advanced stages can become invasive and they can spread to other tissues via lymphatic and/or blood vessels, causing metastasis [Figure 1C].

Table 1. TNM staging system for breast cancer

Primary Tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
Tis (DCIS)	Ductal carcinoma in situ
Tis (LCIS)	Lobular carcinoma in situ
Tis (Paget)	Paget disease of the nipple NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma
T1	Cancer measures 2 cm in diameter
T2	Cancer measures more than 2 cm but not more than 5 cm
T3	Cancer measures more than 5 cm in diameter
T4	Tumor of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules
Regional lymph nodes (N)	
NX	Regional lymph nodes cannot be assessed (eg, previously removed)
N0	No regional lymph node metastasis
Clinical N1	Cancer has spread to lymph nodes under the arm on the same side of breast cancer
Pathologic N1	Cancer is found in one to three lymph nodes
Clinical N2	Cancer has spread to the lymph nodes under the arm, these are attached to each other or to the surrounding tissue, the internal mammary lymph nodes are affected
Pathologic N2	Cancer has spread to the lymph nodes located above or below the clavicle on the same side of the cancer
Clinical N3	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s), with or without level I, II axillary node involvement
Pathologic N3	Cancer has spread to 10 or more lymph nodes under the arm
Metastasis (M)	
M0	There is no cancer spreading
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven > 0.2 mm

Adapted from Sparano et al., 2016.

There are many staging systems, the most widely used for describing the growth and spread of breast cancer is the TNM staging system. The letter T refers to tumor size, N refers to number of lymph nodes to which the tumor has spread and the letter M refers to metastases. Once assigned to the categories T, N and M, this information is combined to designate a general stage of 0, I, II, III or IV (NCCN, 2006) [Table 1].

Other important classification is based on the molecular subtypes. The current classification includes: triple negative, basal-like, Her2, luminal subtype A, luminal subtype B, normal breast like and "claudin-low". They were identified by microarray-based gene expression analysis and unbiased hierarchical clustering. The molecular subtypes display highly significant differences in prediction of overall survival, as well as disease-free survival (Malhotra et al., 2010) [Figure 2].

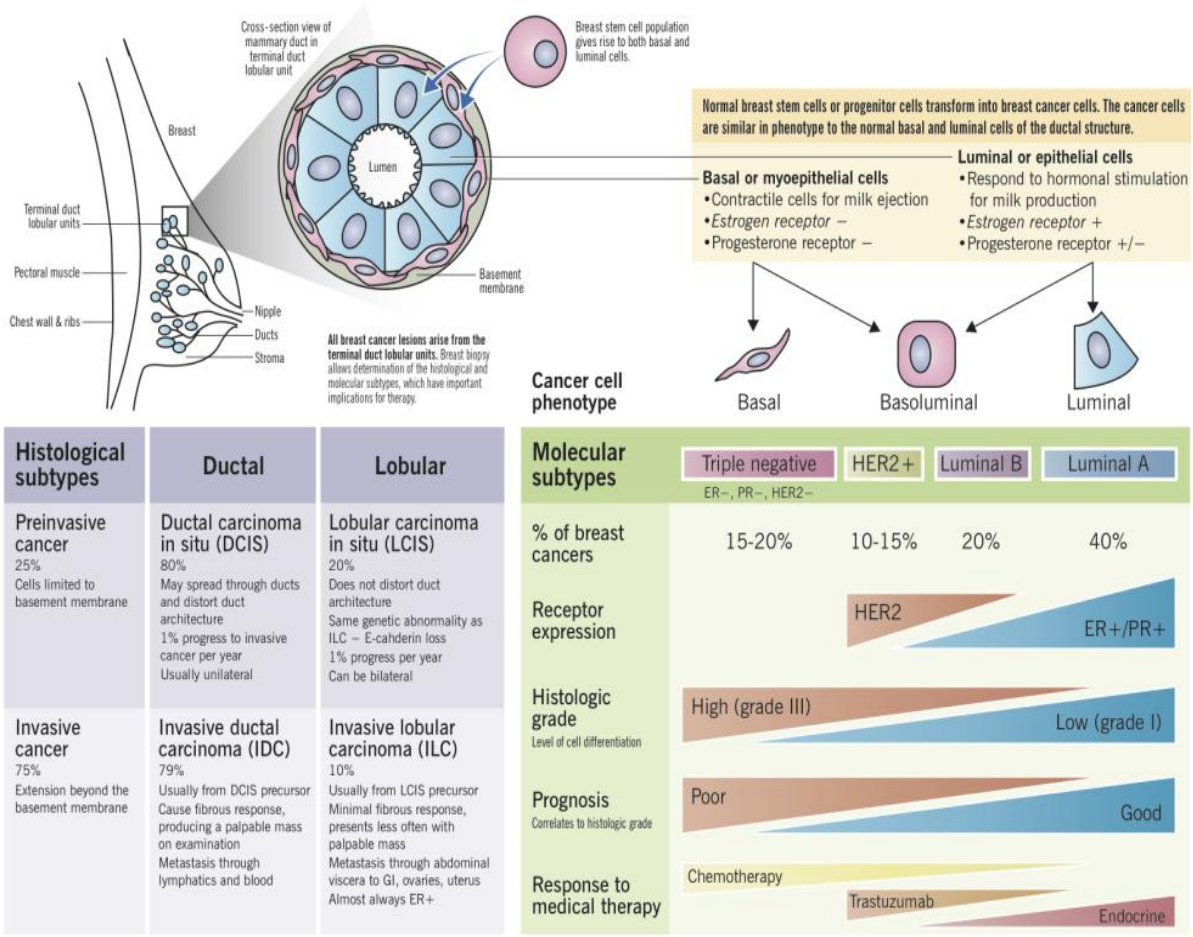


Figure 2. BC pathogenesis and histological and molecular subtypes. General considerations of each molecular subtype are shown. Adapted from Wong et al., 2012.

The luminal molecular subtype is divided in two major categories (luminal A and luminal B), although there is another (luminal C). The luminal A is characterized by the highest expression of

hormone receptors: the ERα (estrogen receptor) gene, GATA binding protein 3, X-box binding protein 1, trefoil factor 3, hepatocyte nuclear factor 3α and estrogen-regulated LIV-1. The subtype B and C showed low to moderate expression of luminal specific genes, including ER. Subtype C is distinguished by high expression of a set of genes like transferring receptor p90, nucleolar protein p40, among others whose function remains unknown, a feature that they share with the basal-like and Her2 subtypes (Sorlie et al., 2001; Cancer Genome Atlas Network, 2012). They respond to endocrine therapy, respond to chemotherapy available (greater in luminal B than luminal A), and show worst prognosis and survival rates in luminal B than luminal A due to poorer tumor grade, larger tumor size and lymph node-positive (Schnitt, 2010).

The basal like subtype, which highly overlaps with triple negative (although not all basal like are triple negative and viceversa), is characterized by high expression of keratins 5/17, laminin, fatty acid binding protein 7, basal epithelial genes and low expression of ER and Her2. A difference between basal like subtype and the triple negative is that the former have expression of certain keratins, whereas the latter typically do not show expression of these markers (Kumar et al., 2015; Sorlie et al., 2001; Cancer Genome Atlas Network, 2012). Clinically, they show ER-/PR-/Her2 - and represent 15% of invasive cancers. They do not show response to endocrine therapy or trastuzumab. Although hypothesized to be potentially sensitive to PARP inhibitors, results from ongoing phase III trials in BC are still awaited (Schnitt, 2010; Robert et al., 2017).

The Her2 subtype is characterized by high expression of several genes in the ERBB2 amplicon at 17q22.24 including *ERBB2* and *GRB7* (Perou et al., 2000). Clinically, they show ER-, PR- and Her2+ markers, represent 15% of invasive cancers, more likely to be high grade and node positive. They respond to trastuzumab (Herceptin) and anthracycline-based chemotherapy (Schnitt, 2010; Goldhirsch et al., 2013).

The normal breast like group shows higher expression of many genes known to be expressed by adipose tissue and other non-epithelial genes (Perou et al., 2000), nonetheless they are less characterized and clinically, they represent a reduced percentage of BC. The other group “claudin low” BC subtype, is characterized by the low to absent expression of luminal differentiation markers, high enrichment for epithelial-to-mesenchymal transition markers and cancer stem cell-like features. Clinically, these latter have poor prognosis, the majority are ER-, PR-, Her2-, triple negative invasive ductal carcinomas with a high frequency of metaplastic and medullary differentiation, that have a response rate amidst basal-like and luminal BC response to standard preoperative chemotherapy (Prat et al., 2010).

2.1.2 Breast Cancer Epidemiology

Breast cancer is the most frequent cancer among women with an estimated 1.67 million new cancer cases diagnosed in 2012, which represents 25% of all cancer cases. Incidence rates vary nearly four-fold across the world regions, ranging from 27 per 100,000 in Middle Africa and Eastern Asia to 92 in Northern America (Ferlay et al., 2015). In addition, incidence rates per country show an increasing trend which could be related with augmented awareness and better screening methods (Torre et al., 2016).

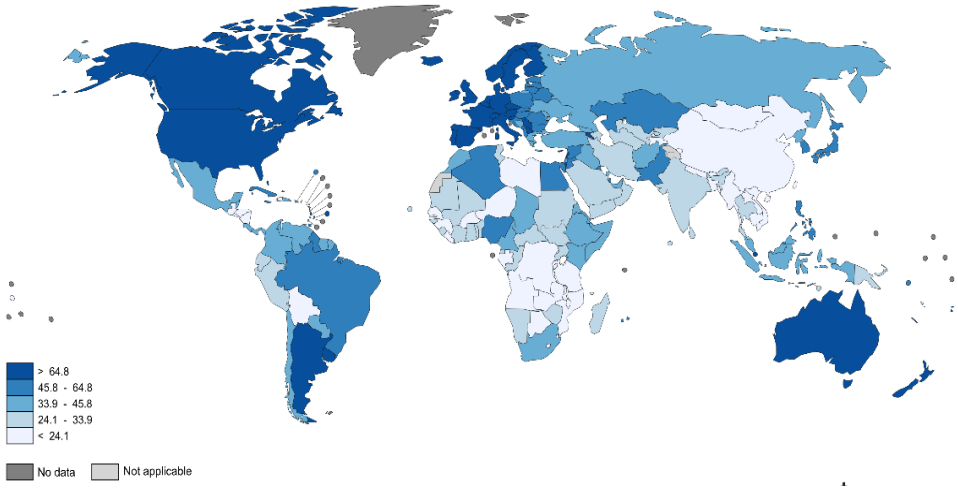
BC ranks as the fifth cause of death from cancer overall (522,000 deaths). Solely in women it represents the 15% of all deaths due to malignant disease, followed by lung and colorectal cancer. The range in mortality rates between world regions is less than that for incidence because of the more favorable survival of breast cancer in (high-incidence) developed regions, with rates ranging from 6 per 100,000 in Eastern Asia to 20 per 100,000 in Western Africa (Siegel et al., 2017). Breast cancer also occurs in men but it is quite infrequent, representing <1% of all the cases of cancer in males, with an incidence estimated as <1 per 100000 men-years (Jemal et al., 2010; Ly et al., 2013).

Recent epidemiology studies show that, e.g. in United States the number of new cases reported in 2016 were 2000 and the number of deaths was 400, this elevated number of deaths are related with advanced state at the time of diagnosis but, in comparison with female cases, the rest of the cases have better prognosis (Siegel et al., 2017) [Figure 3].

2. 1.3 Breast Cancer Risk factors

There are a number of factors that increase the possibility of developing BC. Women at high risk can use this knowledge toward preventive measures, such as: earlier screening, prophylactic surgery and chemoprevention, always with the guidance of their clinician. Some of these factors have been associated with higher (>4.0), moderate (2.1-4.0) or a lesser risk (<2.0) (American Cancer Society, 2015).

A)



B)

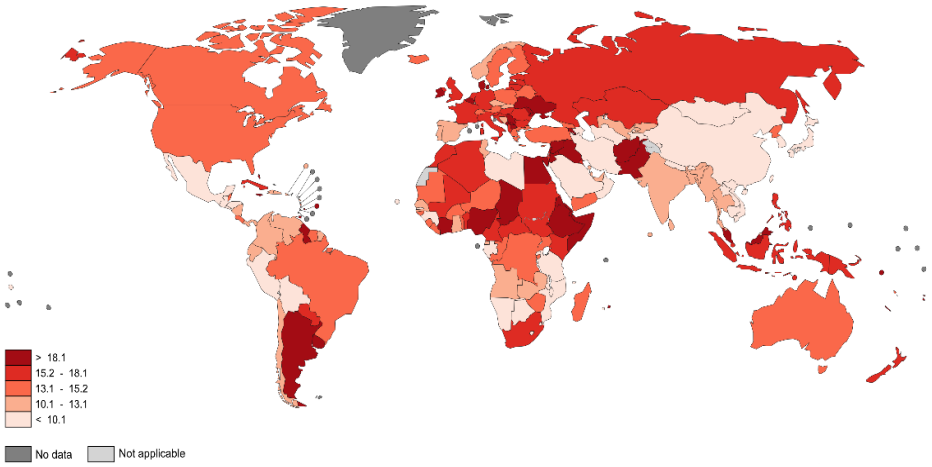


Figure 3. Estimated breast cancer incidence (A) and mortality (B) worldwide. The data represent estimated age-standardized rates per 100,000 women. (Ferlay et al., 2015).

Some of the main BC risk factors explained:

- Gender: being female increases the risk, as this disease is quite infrequent in males, the latter account for a total of 1% of all the breast cancer cases (Speirs et al., 2009).
- Age: the incidence of BC is extremely low before the age of 30, after which, the risk increases linearly until the age of 80 years (Singletary, 2003).
- Hormonal treatment: oral contraceptives of frequent and prolonged use, as well as hormone replacement therapy for more than ten years make women more prone to this type of cancer (Martin et al., 2000).
- Personal history of BC: a woman that has been diagnosed with BC, has increased risk to develop it again, in the same or the other breast.

- Reproductive factors: Menarche at an early age has been linked with a risk level of 10 to 20% of developing BC in comparison with women with menarche at age 14, given the early activation of ovarian cycles. Likewise, there is an increased risk of BC if menopause occurs late which is related with longer lifetime exposure to progesterone/estrogen effect. Early age of motherhood at first pregnancy and parity have been found to decrease the risk of BC development. Still, the effect of pregnancy is different for different types of BC. For triple-negative BC, pregnancy seems to increase risk. Pregnancies exert a protective effect probably due to the terminal differentiation of mammary tissue, although this hypothesis is still under research. Breastfeeding is theorized to protect against BC development (McPherson et al., 2000; Azim et al., 2014).
- Genetic inheritance and family history of BC: about 5-10% of all BC cases have a hereditary component, an inherited mutation in high susceptibility genes as *BRCA1* and *BRCA2*, high susceptibility cancer syndrome genes *(TP53*, *PTEN*, *STK11* and *CDH1*) and/or moderate and low susceptibility genes can be directly related with the appearance of BC, especially if there is any familial background of the disease in previous generations. This topic will be widely explained in the next section as the main focus of present thesis.
- Race/Ethnicity: certain populations have founder mutations that make women more susceptible for the development of BC. An example can be seen in Ashkenazi population, where founder mutations in *BRCA1* 185delAG, 5382insC and *BRCA2* 6174delT have a higher population frequency (2%), which is also related with higher incidence of BC specifically in people with this lineage (Levy-Lahad et al., 1997). In addition, white women are slightly more likely to develop BC than are African-American women, Asian, Hispanic and Native American women have a lower risk of developing and dying from BC (Ferlay et al., 2015).
- Benign diseases of the breast: some diseases such as fibrocystic mastopathy, lesions without proliferative activity (fibroadenomas, hyperplasias, cysts, ductal ectasias and metaplasias), as well as proliferative diseases are predisposing factors for the development of BC. Women who have had any of these conditions should be strictly followed for early detection of any malignant modification (Dupont et al., 1993).
- Women with dense breasts on mammogram have a risk of breast cancer that is 1.2 to 2 times that of women with average breast density (Boyd et al., 1998).
- Diet and body mass: Higher incidence of this disease has been reported in people with an excessive body mass, since the adipose tissue is an important extra-gonadal source of bioavailable estrogens (Singletary, 2003; Hankin et al., 1978).

- Environmental factors: There are many factors that are strongly related to lifestyle and also to geographic distribution. For example, the development of BC is more frequent in urban populations. Regarding lifestyle, BC risk increases if there is no physical activity and due to a sedentary life. Also, there are some evidences linking BC development with the use of products that contain high number of chemicals, like artificial dyes for the hair. Although the exact risk has not been calculated, a link has been established between BC and: alcohol consumption, tobacco use, stress, depression, use of medications such as reserpine or diazepam, spironolactone, exposure to electromagnetic and UV radiation, ionizing or contaminants as well as organochlorined pesticides, low vitamin D levels and unhealthy food intake (Ali et al., 2014; Moukayed et al., 2017).

2.2. Hereditary breast cancer

The majority of BC cases are sporadic (>90%), the rest of the cases, approximate 5-7% are hereditary cases (Melchor et al., 2013). Paul Broca was the first to describe a family with high prevalence of carcinoma of the breast; he made a pedigree of four generations highlighting BC incidence in every generation, which pointed to the fact that BC indeed, could be inherited (Steel et al., 1991). An important proportion of hereditary cases, around 15%, are attributed to germline mutations in either *BRCA 1/2* (breast cancer susceptibility gene 1 and 2). Other genes associated with cancer syndromes, as well as moderate to low susceptibility for BC explain other 20% of the hereditary BC cases (van der Groep et al., 2011) [Figure 4A], but there is still more than 60% of the cases in which the genetic cause is not known, which translates into the impossibility for the patients to benefit from current available genetic tests and so, measurements for prevention, diagnosis and prognosis are based in clinics but not supported by genetic knowledge of the hereditability (Rudolph et al., 2016).

The genetic basis of inherited predisposition to breast cancer has been persistently investigated in the last decades, by using the following techniques: linkage studies (genetic loci are mapped using samples from many members of a large family of an interrogated phenotype to identify genomic segments shared among affected members and absent in healthy members to help localize the area of the genome related with predisposition), resequencing studies (mutational screening of candidate genes, selected in the basis of similarity to known genes associated with a specific phenotype, large numbers of cases and controls are required to identify and compare the total number of pathogenic mutations), Genome Wide Association Studies (GWAS, are studies in which hundreds of thousands of SNPs alleles of haplotypes are tested for a disease in hundreds or thousands of unrelated cases and controls so that association can be inferred as to the location of

the risk variants in the genome) (Aloraifi et al., 2015; Dawn Teare et al., 2005; Manolio, 2010). Major discoveries with these technologies lead to the establishment of three well defined classes of susceptibility alleles for BC: high-penetrance alleles (confer a relative risk of developing BC >4.0 fold higher than the general population risk and are very rare in population), moderate-penetrance alleles (confer a relative risk of developing BC 2.0-4.0 fold higher than the general population risk and are rare in population), and low-penetrance alleles (confer a relative risk of developing BC <2.0 fold higher than the general population risk and more common), which will be described in the next section (Stratton et al., 2008) [Figure 4].

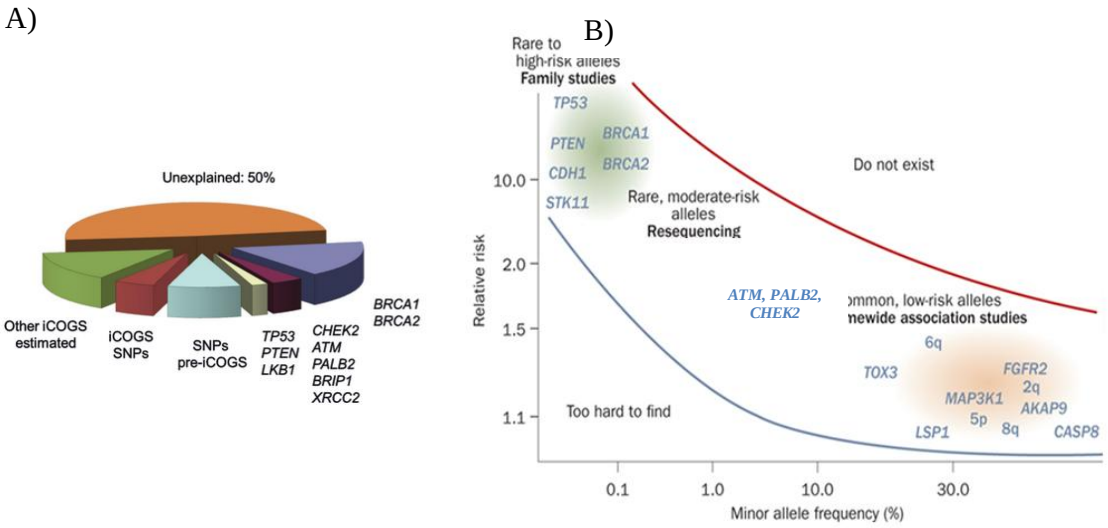


Figure 4. (A) Proportions of familial risk of BC explained by inherited variants. (B) Relative risk in relation to minor allele frequency of high, moderate and low susceptibility genes for breast cancer (Bahcall 2013). The main genetic technique used for the characterization of the different type of genes is marked in bold (Adapted from Rudolph et al., 2016 and Foulkes, 2008).

2.2.1 High susceptibility genes

Only two genes that confer high susceptibility to develop HBOC have been described, *BRCA1* and *BRCA2*, which were discovered by linkage analysis (Miki et al., 1994; Wooster et al., 1995). Germline pathogenic mutations in *BRCA1* confers a lifetime risk of BC of 41-85% for females and 4% for males and in *BRCA2* 50-80% for females and 7% for males. For ovarian cancer, the lifetime risk differs between carriers of these two genes, ranging from 11 to 62% in *BRCA1* mutations carriers and from 10 to 20% in *BRCA2* mutation carriers (Mavaddat et al., 2013). Depending on the population studied, the estimated risk tends to be more specific, e.g. in Spanish population, the average cumulative risk of breast cancer to age 70 years was estimated to be 52%

for *BRCA1* mutation carriers and 47% for *BRCA2* mutation carriers. For ovarian cancer, the estimates were 22% and 18%, respectively (Milne et al., 2008).

The characterization of *BRCA* genes meant a cornerstone in diagnosis, prognosis and treatment for carriers of mutations in these genes. In this regard, genetic testing to identify potential germline pathogenic variants in *BRCA1/2* should be performed to women that present any of these characteristics: two relatives with BC, diagnosis at or before age 35 years old, three or more relatives with BC regardless of age at diagnosis, a combination of both breast and ovarian cancer in relatives, bilateral breast cancer cases in the family, two or more relatives with ovarian cancer regardless of age at diagnosis, history of BC in a male relative, triple-negative BC diagnosed before 60y, Ashkenazi Jewish heritage with any relative presenting breast or ovarian cancer (Bradbury et al., 2007).

Despite all the knowledge gained with *BRCA1/2*, mutations in these genes do not account for more than 20% of all the hereditary cases, which means that there should be other genes that confer high susceptibility that have not been identified (Apostolou et al., 2013). Mutations in other genes such as *TP53*, *PTEN*, *STK11* and *CDH1* confer high susceptibility to breast cancer and are associated to cancer related syndromes: Li-Fraumeni (Gonzalez et al., 2009), Cowden (FitzGerald et al., 1998), Peutz-Jeughers (Hearle et al., 2006) and hereditary diffuse gastric cancer syndrome (Schrader et al., 2008). Lifetime risk for development of BC conferred by germline mutations in these genes are: 50-90%, 50-85%, 32-54% and 52%, respectively. Mutations in these genes also are related with the appearance of other type of tumors like: gastrointestinal, sarcomas, thyroid, renal, stomach, lung and testis cancer (Tan et al., 2012; van Lier et al., 2010).

2.2.2 Moderate susceptibility genes

Mutational screening of candidate genes, led to the discovery of genes whose inherited mutations confer an intermediate increase in risk of breast cancer (Beggs et al., 2009). Other genes that confer moderate susceptibility to HBOC are *CHEK2* (Meijers-Heijboer et al., 2002), *NBS1* (Heikkinen et al., 2006), *BRIP1* (Seal et al., 2006), *PALB2* (Rahman et al., 2007), *RAD51C* (Meindl et al., 2010), *RAD51D* (Loveday et al., 2011) and *ATM* (Renwick et al., 2006). In more recent studies, some of these genes have been established as almost conferring a high-risk for BC rather than moderate risk. Such is the case of *PALB2*, which confers a risk that may overlap with that of *BRCA2* mutation carriers (Antoniou et al., 2014). Other approaches, in contrast, have discarded some genes originally described to confer moderate susceptibility for BC and nowadays additional evidence shows that mutations in genes like *BRIP1* (Ramus et al., 2015; Easton et al.,

2016), *RAD51C* (Song et al., 2015) and *RAD51D* (Thompson et al., 2013) are enriched among ovarian cancer and not breast cancer patients.

It is worth to emphasize that the majority of these genes have been identified through candidate gene screening, selected upon the biological pathways where *BRCA1/BRCA2* were initially implicated, assuming that as *BRCA* genes are implicated in DNA repairing/maintaining systems, then other functional partners encoded by genes with similar function should exist. Another important feature about this group of genes is the fact that they are related with Fanconi Anemia (FA) pathway.

FA is a rare disease with an incidence of 1/350 000 births with a clear recessive pattern of inheritance by biallelic mutations in any of the 21 *FANC* genes described until today. The phenotype is related to abnormalities as short stature, abnormal skin pigmentation, and skeletal malformations of the upper and lower limbs, microcephaly and ophthalmic and genitourinary tract anomalies, bone marrow failure characterized by pancytopenia, thrombocytopenia and leucopenia (Mehta et al., 2002). It has been shown that heterozygous carriers of mutation in FA-genes (e.g. *BRCA2/FANCD1, BRIP1/FANCJ, PALB2/FANCN* and *RAD51C/FANCO*) have increased BC or OC susceptibility (Garcia et al., 2008). Although the mechanisms that link FA and cancer have not been completely explained, the basis arises from the fact that inactivating mutations over them lead to chromosomal instability, accumulation of DNA damage and alteration of major DNA repair pathways: homologous recombination, nucleotide excision repair and mutagenic translesion synthesis which would be intrinsically translated to transformation and malignancy (Moldovan et al., 2009).

2.2.3 Low susceptibility genes

Multiple SNPs that can modify the risk of developing HBOC have been described in the last years and they have been discovered using GWAS and large scale replication studies. Numerous variants have been found so far, which contribute to a small percentage of disease burden. Also, many variants are located either in noncoding or intronic regions and their significance are difficult to interpret (Lynch et al., 2013). Another consideration is that, because of the significantly reduced penetrance and strongly non-Mendelian patterns of inheritance, there is often considerable uncertainty about the exact underlying genetic variant in relation with the low susceptibility genes (Collins et al., 2011). In comparison with high/moderate susceptibility alleles, the low BC susceptibility genes are mostly related with growth-promoting genes as evidence seen for susceptibility variants in *FGFR2* and *FAM84B/c-MYC* (Easton et al., 2007). Though, many

other loci are anonymous or have functions previously unrelated to cancer development which complicates profound understanding of their role in inheritance (Stratton et al., 2008).

Since the publication of the first GWAS studies, there have been major advances regarding low susceptibility alleles. In the first one, five significant loci were described; in the last decade other 78 have been described (Michailidou et al., 2013). These almost 100 common BC risk loci explain 14% of inherited genetic component, with relative risk between 1.05 and 1.26. An important characteristic about recent studies is the pathways being explored to find function relevant genes mapped by GWAS hits: DNA damage recognition and repair, apoptosis, estrogen receptor signaling, tumor progression, metastatic disease and epigenetic changes. In the post-GWAS era a novel combination of strategies is needed to identify target genes and/or function effect of a confirmed loci, these include: fine-scale mapping the locus, bioinformatics predictions for functionality and *in vitro/in vivo* experimental verification of molecular mechanism associated and risk prediction modelling (Fachal et al., 2015). Breast Cancer Association Consortium (BCAC) has worked very assiduously to identify new risk associated variants in large-scale replication study as part of the Collaborative Oncological Gene-Environtment Study (iCOGS). As part of this work they have associated low penetrance BC susceptibility loci with specific breast tumor subtypes (Broeks et al., 2011). An interesting approach that has been proposed in order to identify women at different level of BC risk by the combination of multiple common susceptibility alleles to render a genetic risk stratification effectuated in 77 BC associated SNPs in 66000 cases and controls. This study lead to the estimation of polygenic risk scores (PRS) to stratify BC risk in women without family history and to refine genetic risk in women with a family history of BC and lead the recommendation to improve risk reduction and screening strategies in women with highest PRS scores. In addition, PRS data combined with other risk factors could also guide different management and prevention strategies (Mavaddat et al., 2015). Another challenge would be related to the clarification of the BC phenotype that mutations in particular loci confer. It is likely that in the next years, more loci that confer low susceptibility to BC will be found, although the clinical utility will be limited until it is known in which pathway and manner are these loci implicated in the disease (Freisinger et al., 2008).

2.3. Next Generation Sequencing (NGS)

2.3.1 Brief concept into NGS

Before the introduction of NGS technology (a concept that stands for the parallel sequencing of thousands of DNA fragments at the same time), the main techniques used for identification of high, moderate and low susceptibility alleles had their own limitations. Linkage analysis could not identify the causal mutation of Mendelian disease with extreme phenotypes where the number of

unrelated samples available was small, which translates into weak statistical power. In addition, linkage analysis was not robust enough for disorders with genetic, locus and/or phenotypic heterogeneity, as well, as reduced penetrance and disorders caused by effect of multiple genes (Bush et al.,2010).

Candidate gene analysis is restricted by: prior physiological, biochemical or functional known aspects of possible candidate genes, complex diseases that show a spectrum of phenotypes, incomplete penetrance, a disease that results from the effect of multiple and combined genes where focusing solely in one gene would not explain entirely the clinics, lack in replication of certain studies, among others (Zhu et al., 2007; Tabor et al., 2002). GWAS analysis faces some important issues: insufficient number of cases and controls can lead to false positives, also to prove the biological significance and the association is daunting if coping with a complex disease, the use of GWAS which is mostly applied to look for low penetrance genes not for high-risk genes, presence of variants that can be present in linkage disequilibrium with other one that is the real functional cause are some of the most criticized aspects of GWAS studies (Pearson et al., 2008). Taking all these into account, there are still a large proportion of BC familial cases with unidentified cause. In the last decade, the advent of high throughput technologies for massive sequencing has happened. The most exploited technologies into NGS are Whole Genome Sequencing (looks into the whole genome, including coding and non-coding regions), Whole Exome Sequencing (includes the whole coding regions of the genome), Amplicon-based sequencing (targets a small and specific region of interest for sequencing), RNA sequencing (evaluates RNA and gene expression) and exon capture transcriptome (baits are used to capture portions of the transcriptome) (Simon et al., 2013; Metzker 2010) [Figure 5]. In the next sections, we will focus specifically in Whole Exome Sequencing and its application in hereditary breast cancer context.

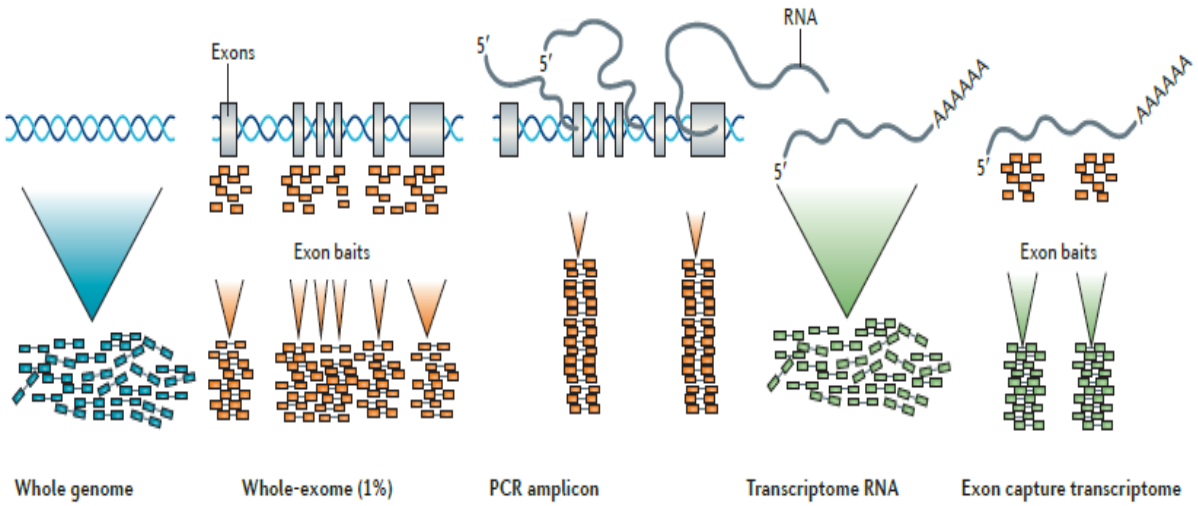


Figure 5. Next Generation Sequencing techniques (Adapted from Simon et al., 2013).

2.3.2 Whole Exome Sequencing in hereditary breast cancer

There have been a number of publications that use WES to study missing hereditability in hereditary breast cancer [Table 2]. The contribution and the effort have been remarkable, but principal findings have not been conclusive as expected and reasons that condition WES scope are exposed in the next section. Nonetheless, in the lasts years a novel putative BC susceptibility gene has been proposed, *RECQL*, meaning that WES can be successful for unraveling genetic causality of HBOC (Cybulski et al., 2015; Sun et al., 2015).

Table 2. Whole Exome Sequencing applied to hereditary breast cancer

Study	Population	Family Selection	Strategy for Candidate Variant Identification	Major Results
Rosa-Rosa et al., 2010 and Rosa-Rosa et al., 2009	Spain	Families selected based in previous study identified by linkage studies candidate regions in Chromosome 3 and 6 (10.8 and 6.5 Mb, respectively) for harboring BC susceptibility genes. Families were BRCAX with at least 2 affected members with BC	Massive parallel sequencing. Analysis pipeline based on SOAP aligner. Hydrid selection on tiling microarrays used for enrichment of exonic sequences within two candidate regions as a valid second step for identification of putative HBOC genes	Not conclusive. 8 candidate SNPs under functional validation with missense, 3'UTR effect and/or interesting gene function. No truncating mutations identified.
Park et al., 2011.	Australia and US	Two greater than third degree affected relatives from four multiple-case, early onset BC families (3). At least 6 cases of BC. Dx under 60y. BRCAX. Negative for other cancer related genes.	Whole exome capture and massive parallel sequencing. Not deeply described	Not conclusive. 2 variants in the DNA damage repair gene *FAN1*. Not causal for BC. Maybe in a familial set and might influence in combination with other risk factors.
Snape et al., 2012.	UK	50 individuals with familial BrCa, bilateral cases (42), age first diagnosis 53, 3 affected per family. BRCAX. 12 selected for validation, no enrichment of group of functionally related genes	Rare, protein truncating mutations and gene function. Excluded read coverage>15, intronic variants, synonymous variants, select protein truncation, with frameshift indel, nonsense mutations and mutations at consensus splicing residues	Not conclusive. Heterozygous truncating variants and 4 mutations in known predisposition genes
Thompson et al., 2012.	Australia and New Zealand	15 high-risk breast cancer BRCAX families with at least four cases of	Quality threshold $\geq$30, Read depth$\geq$10, Allele frequency$\geq$0.15, Deleterious consequences	Not conclusive. Heterozygous, dominant, deleterious mutations in DNA repair genes

		multi-generational BC	nonsense SNVs, frameshift idels, splice variants and complex variants. Priority given in well-established brca-associated DNA repair	*FANCC* and *BLM*, 3 truncating in *FANCC* and 2 in *BLM*
Park et al., 2012	Netherlands, Australia and Spain	Multiple affected relatives from 13 families. BRCAX families. Diagnosis before 45y	Not deeply described	Not conclusive. 2 families with *XRCC2* mutations, one protein truncating and other probably deleterious missense mutation. 6 pathogenic coding variants in 1308 cases.
Hilbers et al., 2013	Netherlands	Six families in which tumours of multiple cases showed a specific genomic profile on array comparative genomic hybridization (aCGH). Gain of almost whole chrs 22.5-6 cases of BrCa per family. Onset 54y. Other cancers. No male cases.	Removed variant with allele frequency>1% in hap map or 1000 genomes, EVS, variant in homozygous state removed. As well, intergenic and non conserved variants in non-coding regions. Priorized truncated protein (gain stop codon, frameshift and splice site variants). Possibly damaging. FunCtion in DNA integrity maintenance.	Not conclusive. Linkage analysis revealed a region initially focusing in the linkage region no potentially pathogenic variant could be identified. Outside of the linkage region multiple variants related to DNA integrity maintenance proteins.
Gracia-Aznarez et al., 2013	France, Italy, Netherlands, Australia and Spain	11 individuals from 7 BRCA1/2 negative families with at least 6 affected women with BrCa diagnosed under the age of 60 across generations.	INDEL and heterozygous SNPs were selected being common to both sequenced members, not being present in 7 HAP MAP controls, not being present in dbSNP130, gene function, DS and QS between 20 and 210. Variants presenting minor allele frequencies below or equal to 1% were reviewed, stop, frameshift and splicing variants were reanalyzed.	Not conclusive. Known variant *CHEK2* 1100del C identified and a catalogue of 11 rare variants associated with BC related with DNA repair, cell proliferation and survival or cell cycle regulation. Support that majority of BRCAX BrCa families might be explained by the action of moderate and/or low penetrance susceptibility alleles
Lynch et al., 2013	US	Family-focused approach. BRCAx family with 5 cases of BC, 4 diagnosed before 50 years old, with mendelian autosomal dominant inheritance.	Calling score >40, mutated base >10 individual sequences, non synonymous codifying changes, sift score 0.05	Not conclusive. Two variants in *KAT6B* in 5/6 BC affected members. No other variants detected in the entire coding exons in 42 BRCAxextra cases.
Kiiski et al., 2014	Finland	24 high risk familial BRCA1/2 negative BC patients (min 3 breast or ovarian cancers in first or second degree relatives) Finnish population	Mean coverage ≥15. Checked in EVS, 1000 genomes, exome data from 144 finish noncancer control cancer. Genes participating in DNA repair were selected. Frameshift deletions and	22 variants of 21 DNA repair genes. Mutation in ***FANCM*** c.5101C>T (p.Q1701X) more frequent o.5% families BRCAX, particular enrichment among patients with triple

			insertion, splicing alteration, missense and nonsense SNVs. Missense predicted pathogenic.	negative BC
Wen et al., 2014	US	Three families with BRCAX familial BC, 17 members with cancer and 5 without cancer. Plus 22 probands for BRCAX familial BC.	Minimum read depth 10 and minimum base quality 30. Known variants in db SNP, 1000 Genomes, Variants causing deleterious effects kept, with Polyphen. Focus in single base change. Removal of variants common in human population, remove family specific normal variants. Selected shared between affected members and functional class of the mutated genes.	Not conclusive. Results indicate that majority of novel deleterious variants identified are family-specific but not shared with other families. In addition to population-approach, family-approach can determine the genetic predisposition. Data cannot be considered as true predisposition without further phenotypic and functional evidences.
Park et al., 2014	US, Australia and Canada	89 women with BC from 47 families with at least 3 cases early onset BC in the family, with bilaterality. Ashkenazi origin and family history	Exome sequencing and multiple parallel sequencing for mutational screening of *RINT1*. Exonic variants plus intronic variants that fell within 20 bp of a splice acceptor 8 bp of a splice donor, allele frequency <0.5%, explored in EVS, 1000G	*RINT1* rare sequence variants associated with intermediate levels of BC risk, similar to the risk conferred by Lynch syndrome
Sun et al., 12015	China	514 familial BC patients, BRCAX and with early onset of BC	Variants priorized were heterozygous which led to truncating mutation or splice-site variants, looking into db SNP and 1000G database	*RECQL* mutations are associated with BC susceptibility
Cybulsky et al., 2015	Poland and Canada	195 women with familial BC, numerous early BC cases, with 3 or more affected relatives, familial story of ovarian cancer.	Truncating variants (insertions, deletions, stop codons) and variants in consensus splice sites. Variants seen in more than two cases and genes with two or more variants. Genes selected upon its relation with cancer pathogenesis	*RECQL* mutations are associated with BC susceptibility
Noh et al., 2015	Korea	A BRCAX family with three siblings, two of them affected with BC, one healthy. No cases of cancer in previous generations	SNVs and indels priorized, exonic or splicing regions, compared in dbSNP and 1000G databases, compared with dbNSFP (functional prediction and annotation of potential non synonymous SNP), Sift score <0.05	Not conclusive. 7 risk variants in genes: *XCR1, DLL1, TH, ACCS, SPPL3, CCNF, SRL*. Genetic evidence should be confirmed by functional analysis.
Määttä et al. 2017	Finland	13 high-risk hereditary breast	Not deeply described	Not conclusive. Variants in *ATM, MYC, PLAU,*

		and/or ovarian cancer families		*RAD51*, and *RRM2B* were enriched in female HBOC patients compared with controls. A rare nonsynonymous variant in RAD50 was detected in a male BC patient. In addition, a very rare BRCA1 variant was identified in a single high-risk family
Jalkh et al., 2017	Lebanon	45 unrelated patients with inherited BC, mean age of diagnosis 44y	Variants selected according to existence in databases (dbSNP), Human Genome Mutation Database HGMD, EVS, 1000genomes, BIC, Leiden, COSMIC, BRCA Exchange website, <1% allele frequency,	Not conclusive. 19 pathogenic mutations in: *ABCC12*, *APC*, *ATM*, *BRCA1*, *BRCA2*, *CDH1*, *ERCC6*, *MSH2*, *POLH*, *PRF1*, *SLX4*, *STK11*, *TP53*. *BRCA1* p.C44F detected twice in the cohort, suggestion a founder effect.
Kim et al., 2017	Egypt	5 BRCA X high risk families.	Variants checked in dbSNP, 1000G, ESP6500. Also in Egyptian genome variation data to filter out polymorphisms, MAF≤1%	Not conclusive. Genetic predisposition for familial BC can be ethnic-specific. Novel variants identified in already known predisposition genes like *ATM*, *BRIP*, *CHEK2*, *RAD51C*, *RAD51D*, *TP53*.

Bold marks, highlight study with successful finding of a potential susceptibility gene in HBOC.

2.3.3 Limitations of WES studies in hereditary breast cancer

Some of the reasons of why the WES studies performed so far in hereditary breast cancer, have not detected another high susceptibility gene imply:

- Model of inheritance: the studies using WES in looking for novel susceptibility gene for HBOC have assumed the existence of other(s) high susceptibility gene(s) that follow the inheritance model of *BRCA1/2*, which is the dominant model. Some publications have implemented mathematical simulations in a large numbers of families in which at least a percentage of the families with unknown genetic basis could be associated with a recessive model which has not been explored. In a three generation population-based families Australian study, in non-carriers of *BRCA1/2* mutations, about 1/250 (95% confidence interval) women have a recessive risk of 86% of early breast cancer by age 50 years and near certainty by age 60 years (Cui et al., 2001). Meanwhile, in another study including 2531 Ashkenazi Jewish families, Mendelian transmission of a major recessive gene (s) was compatible with the data, predicting that 4% of women would carry high-risk genotype and 85% of them developing BC at 70 years old (Kaufman et al., 2003).

- Incomplete design of the study: This refers, for example, to partial inclusion of the members of a family, due to the impossibility to gather all required samples. A small number of samples limit the statistical power of inference for risk assertion. An extra effort has to be put to the selection of families with extreme phenotypes to increase the possibility of finding something meaningful among the noise (Petersen et al., 2017).

- Penetrance of the mutations: germline mutations are not fully penetrant, this means that the mutation could segregate into different members of the family but not all the carriers develop the disease (Ott et al., 2015).

- Polygenic model: it seems that there is a percentage of hereditary BC cases where the genetic cause of the disease is not solely a gene, rather, it is the result of susceptibility alleles in many different genes in which, large numbers of susceptibility polymorphisms act multiplicatively on risk, modifying it. Although this hypothesis is difficult to prove in all the families in which there is not known mutation in major high/moderate susceptibility alleles detected, there is statistical data that supports high polygenic risk scores for risk prediction in hereditary BC families (Antoniou et al., 2003; Antoniou et al., 2006; Pharoah et al., 2002; Muranen et al., 2016).

- Presence of variants of unknown significance (VUS): from every NGS study, there are a lot of variants whose implication is estimated *in silico*; although for many of them there is no functional study that demonstrates conclusively its role in the disease (Rainville et al., 2014). Efforts through different consortia are being promoted to clarify significance of VUS in HBOC, as well as setting guidelines for the classification and interpretation of sequence variants (COMPLEXO et al., 2013; Richards et al., 2015).

- Phenocopies: it refers to a member of a family that lacks specific inherited genotype, although, in the context of familial breast cancer, develops the disease. Presences of this type of events difficult the use of WES to study a hypothesized model of inheritance (Smith et al., 2007).

- Other considerations: there is the possibility of an unusual cluster of environmentally caused (non-familial/genetic) cases of the disease which would be taught at first glance as familial, but with genetic test would be unmasked. Also, careful considerations have to be taken when the pipeline analysis is designed so that interesting variants are not ruled out by filters since the beginning (Feng et al., 2011).

2.3.4 Clinical Relevance

There have been attempts to translate current genetic knowledge in uses beyond classical genetic counseling of genes that have been included into screening associated to breast cancer risk, since the discovery of *BRCA1/2*. Application of WES in oncology is set toward identification of somatic mutations, germline mutation, novel driver mutations, mutation network reconstruction and identification of predisposing variants (Rabbani et al., 2014; Xuan et al., 2013).

Multiple gene-sequencing is an emerging field in the eye of the scope. Nonetheless, clinical application is daunting, due to incorrect inferences, potential inappropriate clinical application and the overwhelming amount of data for which evidences of its implication as the causality of the disease is still missing. The management of variants of unknown significance is also a very delicate issue in both clinics and research (Rahman 2014). Nevertheless, the objective is set toward the implementation of NGS panels to identify pathogenic mutations in BRCAX families without a known genetic cause (Graffeo et al., 2017). This is extremely important in the set of management and prevention of both patients affected and carriers of the mutations. Different studies have already informed about the value of this technology for risk assessment in patients with early-onset of familial BC (Nielsen et al., 2016; Bertier et al., 2016; Lin et al., 2016) [Figure 6].

Albeit the interpretation is daunting, NGS has led to the discovery of new predisposition genes, either family or ethnic specific. It is likely that other genes that confer high or moderate susceptibility appear, which a priority is in early diagnosis and prevention, to monitor mutation carriers, halt tumoral development through chemoprevention and applying preventive surgery to breast and/or ovarian tissues (Kim et al., 2017).

In the case of identification of a novel variant in moderate-susceptibility gene by NGS, medical counsel was typically based on the extrapolation of guidelines used for management of individuals with high penetrance variants of cancer susceptibility genes that could translate in significant harm and unnecessary preventive measurements. This is why a framework for clinical decision-making for individuals with inherited moderate penetrance gene mutations associated with an increased risk of cancer has been published. In this, women carrying mutations in these genes could be considered for early mammographic screening and/or breast magnetic resonance imaging at annual starting age of 40 when mutations are present in *ATM*, *CHEK2* (truncating type), *NBN*; whereas in *PALB2*, the annual starting age should be 30 years. And for mutations in *BRIP1*, *RAD51C* and *RAD51D*, the family story has to be considered if the cumulative life time risk (by 80 years) is >20-25%. For mutations in *ATM*, *CHEK2*, *NBN* and *PALB2*, decision-

making over risk-reducing oophorectomy regarding prevention of ovarian cancer should be guided by family history of the disease, if present, because risk-reducing oophorectomy (RRSO) is not indicated based on the presence of moderate penetrance mutations alone. For mutations in *BRIP1*, *RAD51C* and *RAD51D*, consideration of RRSO might be warranted for individuals with mutations in any of these genes if the individual has a clear family history of ovarian cancer (>1 case), especially in close relatives (Tung et al., 2016; Gradishar et al., 2017-NCCN guidelines).

One of the great advantages of finding the causal variant in a BRCAX family is that it can guide personalized treatment for the patient. An example can be seen with the promising results achieved with PARP inhibitors (Farmer et al., 2005). Olaparib is the only PARP inhibitor approved by FDA in the treatment of patients with germline *BRCA1/2* mutated advanced ovarian cancer pretreated with ≥3 prior lines of chemotherapy and other inhibitors like rucaparib with FDA breakthrough in April 2015 (Parkes et al., 2016). This selective specificity in BRCAness condition, depend in the high reliance of the tumoral tissue on alternative DNA repairing pathways, where poly(ADP-ribose) (PARP) polymerase is pivotal for survival, hence its inhibition favors cell death, the so-called synthetic lethality (Herceg et al., 2001; Lord et al., 2017). Being capable of selecting personalized medicine according to specific germline mutation has given crucial alternatives to patients, herein resides the importance of the NGS technologies and their debated implementation in clinics.

In conclusion, despite technical and ethical constraints, WES has lead to the discover of many previously unknown mendelian causes in monogenic disorders. Its use in complex diseases is growing and with better strategies and refinement of bioinformatics' analysis, WES promises to be more conclusive and a determinant factor toward finding missing hereditability in BC.

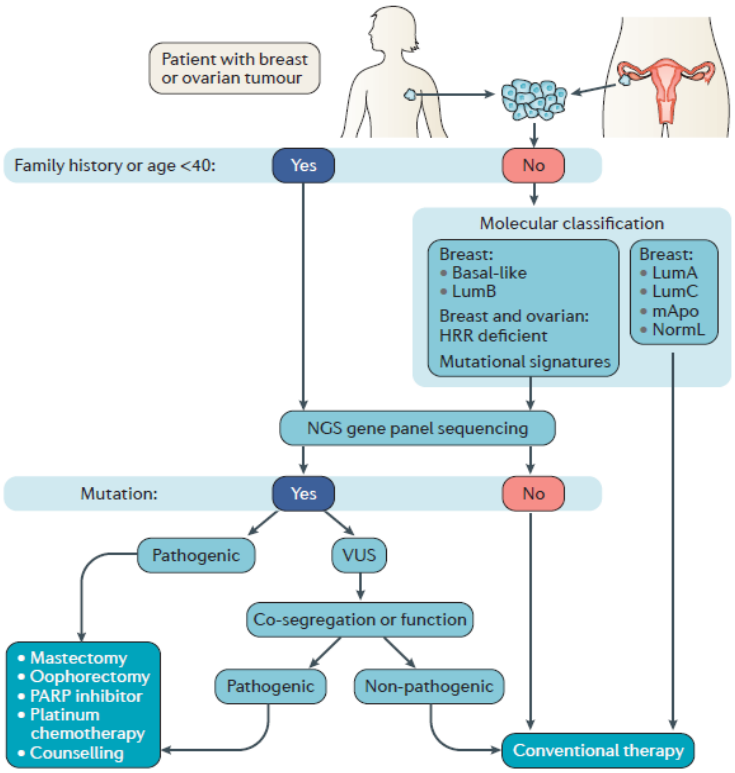


Figure 6. Proposed clinical management of breast and ovarian tumors using NGS techniques. NGS panels would be used to identify germline mutations. In the case of being pathogenic, patients are treated and counseled according to current HBOC recommendations. In the case of identification of a variant of unknown significance (VUS), co-segregation or functional analysis may be undertaken to classify the VUS and if available, guide personalized medicine toward particular treatments (Adapted from Nielsen et al., 2016).

3. General Objective

- Search for new susceptibility genes in hereditary breast cancer families (BRCAX) with an apparent recessive model of inheritance using WES.

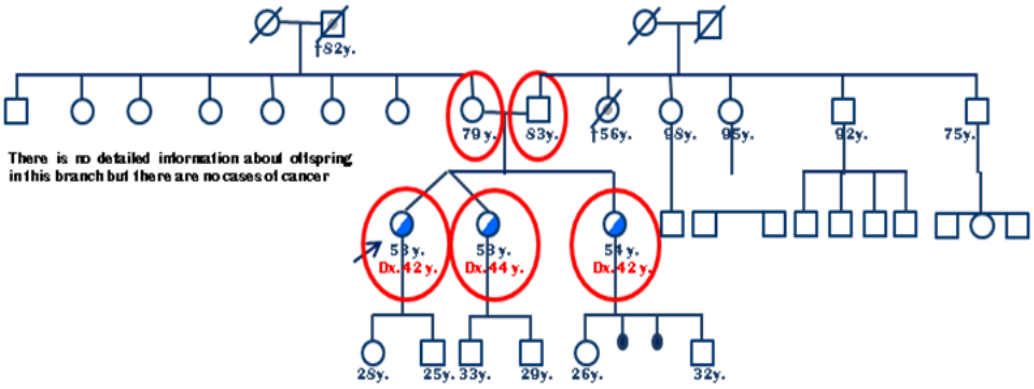
4. Patients and Methods

4.1 Patients and Families of Study

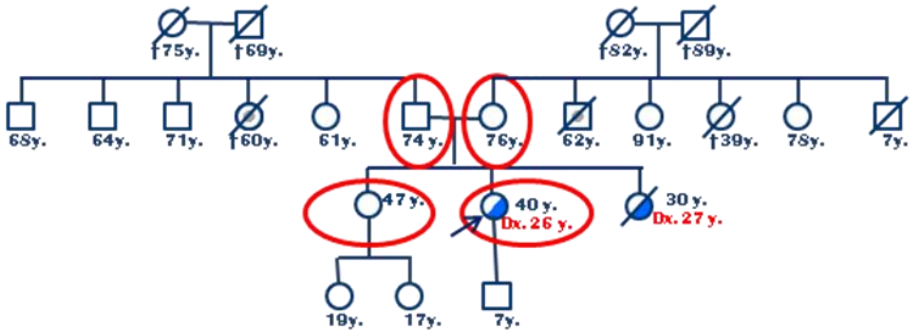
Families of study

Four families with an apparent monogenic recessive model of inheritance for BC where selected based on: presence of two or three siblings affected with BC, early onset of BC (average age under 40 years for the first diagnosis of BC) and absence of familial antecedents of breast or ovarian cancer [Figure 7]. Deleterious mutations in *BRCA1/2* had been previously ruled out by massive sequencing based in the Amplification kit BRCA MASTR V2.1 (Multiplicom) with the GS Junior 454 Sequencing Roche Machine. Blood samples from other members of the family were obtained for further segregation analysis. All participants signed an informed consent approved by the Ethics Committee of Carlos III Institute of Health.

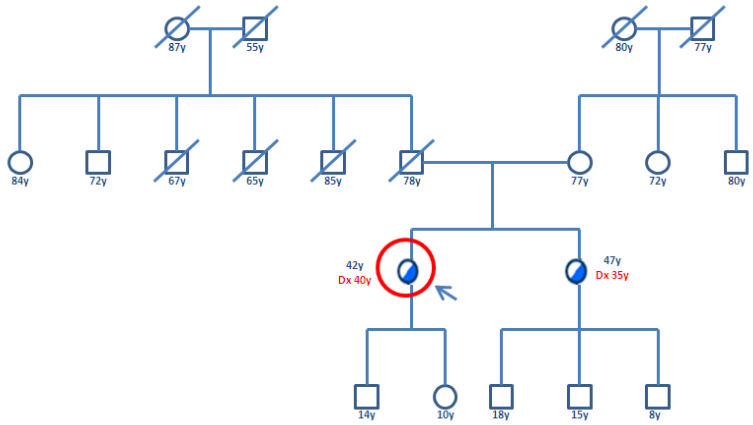
1)



2)



3)



4)

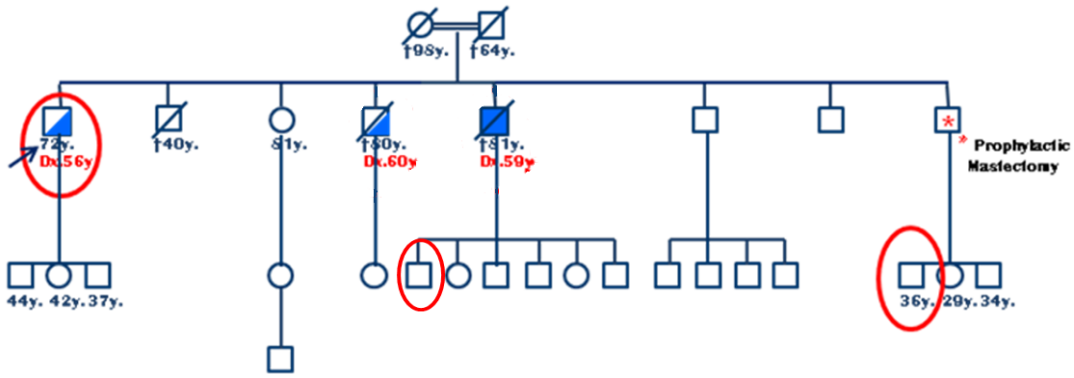


Figure 7. Families with apparent recessive model of inheritance selected for study (1-4). Affected members with unilateral BC are represented with half blue colored circles/squares, bilateral BC are represented with full blue colored circles/squares; Dx, Age of diagnosis; arrow, index case; red circle, analyzed by WES. Other types of cancer are represented with grey dots inside circles.

Cohorts of BRCAX cases and controls used in the different studies that compose this thesis are presented herein:

BRCAX cases (OpenArray study)

To evaluate the frequency of the selected variants found through WES, we performed a case-control association study in 1477 BRCAX families and 589 controls. Index cases from 1477 Spanish breast and/or ovarian cancer families recruited in the Spanish National Cancer Centre (CNIO) (n=577) and Hospital Clínico San Carlos (HCSC) (n=900) were selected for screening of the selected variants. BRCAX families fulfilled one of the following criteria: presented one BC at age <35 or at least two first-degree relatives diagnosed with BC at least one of them at age 50 or younger or fulfilled the same criteria but included at least one ovarian cancer case or at least one case of male BC. All the families were negative for mutations in *BRCA1* and *BRCA2*.

Controls (OpenArray study)

DNA samples from 589 healthy women were collected from the Spanish College of Lawyers (n=271) and Hospital Clínico San Carlos (HCSC) (n=318) between 25 and 65 years of age and without personal or familial antecedents of any type of cancer.

BRCAX families (*ATM* study)

392 families were analyzed for a comprehensive *ATM* mutational study and fulfilled one of the following criteria: presented one BC at age <35 or at least two first-degree relatives diagnosed with BC at least one of them at age 50 or younger or fulfilled the same criteria but included at least one ovarian cancer case or at least one case of male BC. From the 392 families 172 proceeded from CNIO and 220 from Sistemas Genómicos. All were negative for mutations in *BRCA1* and *BRCA2* (BRCAX) and the CNIO families (44%) were also negative for mutations in the *PALB2, RAD51C* and *RAD51D* genes.

Controls (*ATM* study)

350 index cases from families affected with different diseases not related to breast cancer were selected as a comparative group for the comprehensive analysis of the whole coding sequence of the ATM gene. 194 cases were affected with neurological disorders and 156 were affected with different cancer syndromes none of them related to breast cancer. Individuals proceeded from CNIO (59) and Sistemas Genómicos (291), were aged between 18 and 85 and 144 of them were women.

BRCAx families (*RECQL5* study)

Index cases from 700 Spanish breast only cancer families recruited in the Spanish National Cancer Centre (CNIO, n=374) and Hospital Clínico San Carlos (HCSC, n=326) were selected for complete coding sequencing of the gene. BRCAX families fulfilled one of the following criteria: presented one BC at age <35 or at least two first-degree relatives diagnosed with BC at least one of them at age 50 or younger. All were negative for mutations in *BRCA1* and *BRCA2* (BRCAX). Part of these cases belonged to the cohort used for the case-control association study, selected upon the presence of breast only cases.

Controls (*RECQL5* study)

DNA samples from 588 women were collected from the Spanish College of Lawyers (n=271) and Hospital Clínico San Carlos (HCSC) (n=317) between 25 and 65 years of age and without personal or familial antecedents of any type of cancer. Additional 166 control women (from the Spanish National Cancer Centre, CNIO) included in the full coding sequencing study of the gene

had the same characteristics as mentioned. Parts of these controls belonged to the cohort used for the case-control association study.

Male Breast Cancer (MBC) cases and controls

An initial cohort of MBC cases from Spanish population (n=50) was evaluated for the presence of selected variants obtained from the analysis of family 4. The characteristics for the selection of this preliminary cohort were: presence of one case of male breast cancer and negative for mutations in *BRCA1/2*. Importantly, this cohort is enriched in MBC cases that belong to BRCAX families with antecedents of the disease.

A larger cohort of 1200 MBC was evaluated in collaboration with the "Complex Trait Genetics" laboratory, from the Institute of Cancer Research in London, UK. Because of the rarity of the disease, these collection of cases were not selected specifically for familial antecedents, just for personal story of the disease (n=1200). The mean age of diagnosis was 63.8 years, with ranges of age between 23 and 87 years old. Reported MBC bilaterality in only 2% of the cases.

DNA samples from 500 controls (healthy men without personal/familial antecedents of the disease) were used for comparison in the genotyping assay. Mean age of the cohort was 63 years old and age ranges between 23 and 90 years old.

4.2 Methods

4.2.1Whole Exome Sequencing (WES)

Genomic DNA was isolated from peripheral blood lymphocytes using the MagNA Pure LC Total Nucleic Acid Extraction. DNA concentration was determined using PicoGreen dsDNA quantification reagent (Invitrogen). As a concept, in sample preparation, genomic DNA was disengaged, the fragment ends were repaired by T4 DNA ligase, 3' A-tailing was performed followed by ligation of paired-end adaptor to the fragments to finally amplify the prepared library (Rabbani et al., 2012). To enrich the library, hybridization was done with a biotinylated oligo library or DNA baits and captured with streptavidin beads (Exome-seq Nimblegen kit, in this case). Recovery of the hybridized fragments was followed by amplification. Before sequencing, library integrity was verified with highly sensitive methods. Samples were paired-end sequenced on a HiSeq2000 Illumina platform, using two lanes per sample and generating 101 base pair long reads with a commitment >50x.

The amplification/sequencing method employed by Illumina technology was bridge amplification where the enriched library is tethered to a flow cell surface by ligating adapters to both ends of the fragments. Then, solid-phase bridge amplification was performed, resulting in the formation of thousand to millions of clusters from all the fragments of the library amplified. Cycles of sequencing were carried out by adding four fluorescent tagged reversible terminators, which are laser excited leading to the identification of the base being incorporated in every sequencing cycle, which occurred in a massive parallel form for each cluster. As nucleotides incorporated into the growing DNA strand, they were digitally recorded as reads (sequence) which would require further bioinformatics analysis to select the potential variant(s) implicated in the disease (Grada et al., 2013; Chong et al., 2015).

4.2.2.Bioinformatic analysis and variant filtration

Sequencing and processing of the raw data were performed in the National Center for Genomic Analysis (CNAG) in Barcelona, Spain. Briefly, the output format from raw data was given as QC format due to the read quality control usually performed with FastQC program. Preprocessing was used for removal of bad quality reads, adapter reads and trimming of the reads under certain threshold before alignment. Mapping consisted in the comparison of the exome of the sample against the genome of reference, (GRCh37/hg19) using GEM mapper (Kamps et al., 2017). Post-alignment process consisted of read duplicate removal, indel realignment and base quality score recalibration (Bao et al., 2014; Nielsen et al., 2011). Variant calling, annotation and priorization refered to the methods and bioinformatic software used to detect SNVs (single nucleotide variants), indels, CNVs (copy number variants), and large SVs in the sample when

compared with a genome of reference. Realignment and search for indels was performed using GATK. Variant calling was performed with Samtools19 (http://samtools.sourceforge.net). SnpEff (http:/snpeff.sourceforge.net) and snpSIFT were used to filter and manipulate annotated files. Variant filtering was performed in the Spanish National Cancer Research Centre (CNIO). Concisely, variants were prioritized based on sequence quality (Read Depth >10 and genotype Quality >25), impact on the protein (missense, stop lost/gain, frameshift, splice-site) and gMAF value $\leq$ 0.001 for a dominant model of inheritance and $\leq$ 0.05 for the recessive model of inheritance [Figure 8].

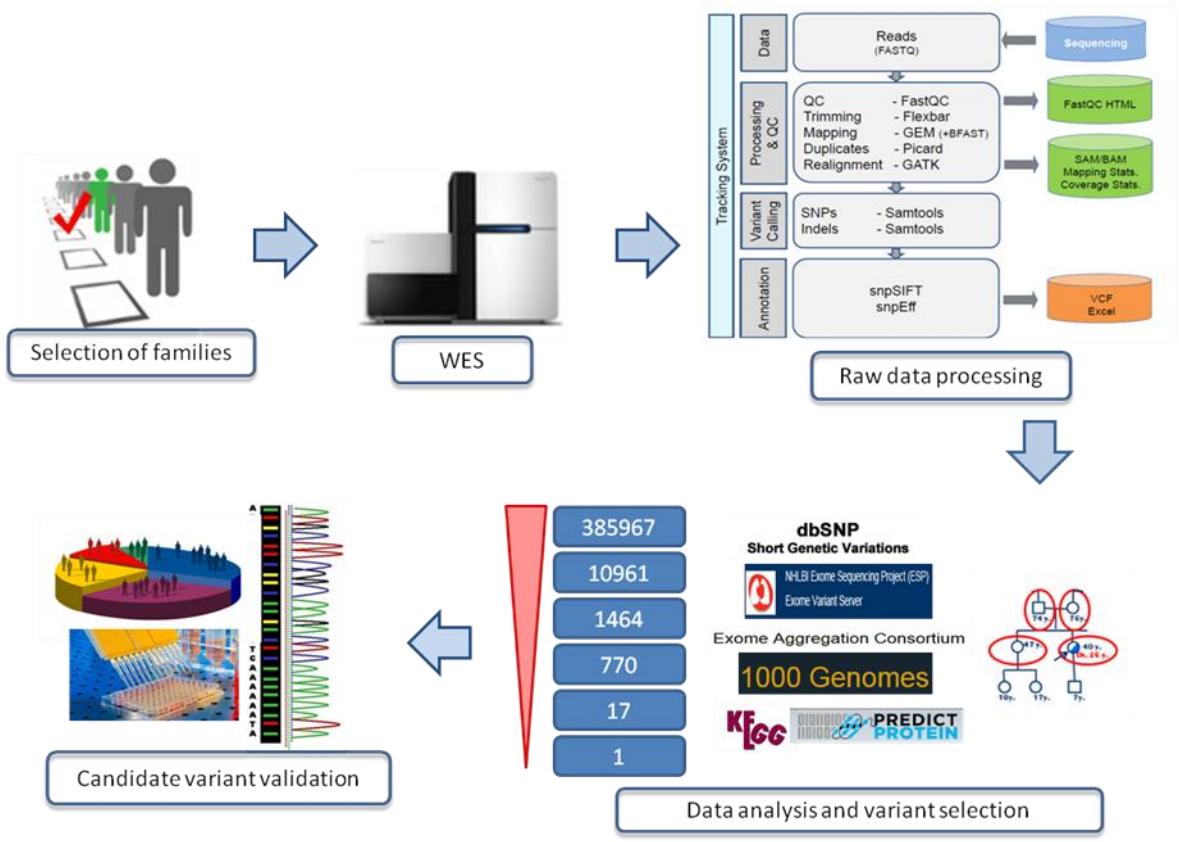


Figure 8. General workflow including bioinformatics pipeline for raw data analysis. After selection of the families with the phenotype of interest, WES and raw data processing was performed according to their in-house pipeline. Data analysis, variant filtering and selection per family were done as explained. An example is given for family 2, total number of variants detected by WES in four family members was 385967, according to the recessive model of inheritance, just homozygous variants were taken into account and variants shared among healthy and affected BC sister were ruled out leaving 10961 variants. Through GMAF number of variants decreased to 1464. Considering quality parameters, 770 variants passed the threshold. Applying impact effect priorization, functional pathways and predictors, we had one variant selected for this family. After that, other approaches applied to assess the implication of the variant in BC were: validation through Sanger sequencing; effect in splicing if any; population studies (mainly, case-control association studies) and functional studies (designed depending in the nature of the gene found).

4.2.3.Segregation analysis and Sanger validation

To verify the status of the selected variants in each family where DNA was available, targeted regions for variants of interest were amplified by PCR technology (Thermo Scientific Maxima Hot Start Taq DNA Polymerase) using a suitable primer pair [Supplementary Table 2]. The products were sequenced using the Big Dye Terminator Cycle sequencing kit on an ABI 3730xl Sequencer (Applied Biosystems). The conditions of the PCR were: initial denaturalization 94ºC per 5 minutes, followed by 35 cycles of three steps (denaturing at 94ºC per 30 seconds, hybridization at customed temperature set by a previous PCR in gradient per 30 seconds and extension at 72ºC per 45 seconds), followed by a final elongation at 72ºC per 7 minutes. Sanger validation was used to confirm the status of the variant in the family as well as to validate the variants detected through WES. On average, a 80% of the variants was validated through Sanger sequencing.

4.2.4 Case-Control Association Study

Case control association study was performed with OpenArray Real-Time PCR technology (Applied Biosciences), according to manufacturer's standard protocol. This platform is based on Taqman a genotyping assay that allows parallel genotyping of 64 SNPs in more than 2000 samples at the same time and allows the detection of a studied variant in a large number of controls and cases to establish association with the interrogated disease (Roberts et al., 2009). Whether probes could not be evaluated due to technical limitations in the OpenArray or due to the design of the probe for this technology, normal Taqman assays were designed and performed for the case-control association study of specific SNPs. HapMap, sample duplicates and samples with status of the variant confirmed by Sanger sequencing were included in the array to serve as internal controls. Genotype calling and sample clustering was performed in TaqMan Genotyper Software v1.3 (Applied Biosciences).

4.2.5. Splicing studies

Variants predicted to trigger aberrant splicing were tested in RNA extracted from blood samples from affected members to verify its pathogenicity. RNA was isolated from peripheral blood lymphocytes from using TRIZOL reagent according to manufacturer instructions. RNA was quantified and samples with a ratio A260/280 ratio>1.8 were taken into account. cDNA synthesis was performed using 500 ng of RNA and the High Capacity cDNA reverse transcriptase kit (Applied Biosystems). Primers were designed to test found variants [Supplementary Table 2] and evaluate its impact.

4.2.6. Loss of heterozigosity (LOH) analysis

When possible, paraffin-embedded tumor from affected members of the families was obtained. Morphologic diagnosis was made in 4u hematoxylin and eosin (HE)-stained sections of tumor, after 10% formaldehyde fixative (24 hr) and paraffin embedding. Case was graded with a modified Bloom-Richardson score. The proportion of tumor cells was rated over 85 - 90%. DNA was extracted using DNA easy Blood and Tissue kit (Qiagen, Chatsworth, CA). PCR amplification was performed with a new set of primers for an amplicon of 200pb and Sanger sequencing was performed in order to compare with genomic DNA from the patient and detect LOH.

4.2.7. Full coding sequencing of *RECQL5* using the Illumina TruSeq Custom Amplicon platform

Analysis of the whole coding sequence of *RECQL5* gene was performed using Illumina Truseq Custom Amplicon designed specifically for the gene of interest adapted with 384 indexes.

Briefly, 50 ng of the index case gDNA was reconstituted. Custom oligo pool, was hybridized by adding 5ul of a dilution of RS1with Custom Amplicon Oligo Tube (CAT) in a ratio 1:2, with 15 ul of OHS2 (Oligo hybridization for Sequencing) per sample. After hybridization, several washes steps were performed to ensure removal of unbounds oligos. Then, a DNA polymerase extended from the upstream oligos through the targeted region, followed by ligation to the 5' end of the downstream oligo using a DNA ligase. Products containing the targeted regions of interest flanked by sequences required for amplification were formed. The conditions of the PCR were: 37ºC for 45 minutes, 70ºC for 20 minutes and hold at 4ºC. Libraries were then amplified, adapters added as well as sequences required for cluster formation. The conditions of the PCR adjusted for less than 96 amplicons were: 95ºC for 3 minutes, 32 cycles of 98ºC for 20 seconds, 67ºC for 20 seconds and finally, 72ºC for 40 seconds.

Libraries were cleaned with sample purification beads to purify the PCR products from other reaction components. The correct amplification of the libraries was confirmed in this step by running 2 ul per sample in an agarose gel 2%. The expected PCR product size for these libraries was around 350 bp. At a final stage of the Truseq protocol, libraries were normalized for a balanced representation. After pooling the libraries (combine equal volumes of normalized libraries), these were diluted with HT1 according to manufacturer protocol and denatured by letting stand the samples at 96ºC per 2 minutes and 4ºC per 5 minutes. To verify the formation of the clusters, 30 ul of phiX 12.5 pM, was used as an internal control. Finally, libraries were further ultrasequenced in a single run in a MiSeq H2000 platform from Illumina. Five runs in total.

Pipeline analysis of raw data was performed using VariantStudio 2.0 Software included in the platform of Illumina. The settings for priorization of the variants were: heterozygous state, prioritizing missense, frameshift and/or stop gain/lost variants, with an alternative variant frequency >10% of the reads and a population frequency <1% in European population. Also, variants presenting low read quality and present in more than 20 heterozygous (per allele count) in Exome Aggregation Consortium (EXAC) (http://exac.broadinstitute.org/, last accessed March 2017) were ruled out as putative deleterious. Clearly pathogenic germline variants were selected upon the lost/gain of a stop codon or effect of a frameshift mutation that generate a truncated product. Potential germline pathogenic variants were selected based on the combination of eight *in silico* predictors SIFT, MUTASTER, Polyphen-2, FATHMM, SNP&Go, Mutation Assessor, MUTPRED and Condel, where at least 5 of them indicated pathogenicity. In addition to the number of predictors, Predict Protein Score (PPS, calculus by position) (https://www.predictprotein.org/) was taken into account to consider pathogenic/tolerated effect. A score higher than >50 was considered as a high pathogenic predicted effect. Variants affecting intron boundaries (+/-5 pb) and predicted to affect nearest splice site by 5 in silico splicing predictors (MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) were also taken as potentially deleterious. Missense variants that laid in the end of the exons were considered although they did not pass the filters of the predictors, as they could induce aberrant splicing. All clearly/potential deleterious candidate variants for *RECQL5* gene were confirmed by Sanger sequencing in index cases samples [Supplementary Table 3].

4.2.8. *In silico* inference of missense variants effect in RECQL5 domains

In silico studies were performed for putative proteins encompassing the selected mutations in comparison with the putative wild type (wt) RECQL5 protein. 23 variants were analyzed including 9 in which the helicase activity had been tested experimentally (Newman et al, 2017), the 10 variants putatively predicted as deleterious from our case-control study and 4 variants considered as negative control: 1 with high MAF (c.1439A>G; p.Asp480Gly (rs820196) quite frequent in our cohort and a GMAF in EXAC of 0.2 in general population) and 3 with neutral effect based on damage predictors). (ATP)M and (ENZIME CONC) values were taken from Newman et al., 2017. Putative proteins encompassing mutations in D2 domain were considered highly deleterious while putative proteins encompassing mutations in D1 and Zn^{2+} had intermediate effect. Neutral effect was considered for putative proteins with mutations annotated after p.435 position.

Solvent accessibility notation (PACC) was calculated for all positions of the putative wt RECQL5 protein and compared to PACC of putative proteins encompassing mutations from Newman study. Solvent accessibility of amino acids (PACC score) was predicted using the

PROFAcc algorithm (Schlessinger et al., 2005) implemented in ProteinPredict (Yachdav et al., 2014). Only one position (p.283) changed PACC score in putative proteins from Newman et al., 2017, compared wt putative protein. PACC score for this position was calculated for all putative proteins encompassing the studied mutations. Secondary structures (β-strands, α-helix and loops) of the putative protein models were based on REPROFSec prediction. The predictions of the annotation (minimum REPROFSEc score of 5) of conserved secondary structures for wt RECQL5 were based on several original prediction methods (NORSnet, DISOPRED2, PROFbval and Ucon) implemented in PredictProtein (Yachdav et al., 2014). Wt RECQL5 annotation was compared with putative proteins encompassing the studied mutations to determine loss/gain of secondary structures.

Protein binding regions were found using the ISIS algorithm (Ofran et al., 2007) and SomeNA predictor method (Hönigschmid, 2012). Putative Protein-binding sites score (threshold >20) were calculated for wt RECQL5 protein. Selected positions at protein-binding domains were evaluated in RECQL5 mutants. Most significant score changes (threshold <15) were considered highly deleterious and threshold between 15>20 were considered intermediate effect.

The clustering score was assigned regarding relative effect for each in *silico* study and normalized by the total number of *in silico* studies performed for each putative protein. Clustering was performed by score sorting and only considering the *in silico* studies included in the domain where the mutation was annotated.

4.2.9. Mutational Analysis of *ATM* gene

Analysis of the whole coding sequence of the *ATM* gene was performed using Onco-Gene SGKit LV2511 according to manufacturer's protocol. Concisely, 50 ng of index case gDNA was tagmented, amplified and libraries were purified. Quality and quantity assessment was performed on an Agilent 2100 Bioanalyzer instrument. Libraries were then hybridized with capture probes for targeted regions. Capture system consisted in streptavidin coated magnetic beads to select probes hybridized. Libraries were indexed for combination in a single run for further ultrasequencing in a MiSeq platform from Illumina.

Familial cancer-relevant genes analyzed in the kit englobe: *APC, ATM, AXIN2,* ***BARD1****, BLM, BMPR1A, BRCA1, BRCA2,* ***BRIP1****, BUB1, CDH1, CDKN2A,* ***CHEK2****, DDB2, DKC1, ELANE, EPCAM, ERCC1 , ERCC2, ERCC3, ERCC4, ERCC5,* ***FANCA****,* ***FANCB****,* ***FANCC****,* ***FANCD2****,* ***FANCE****,* ***FANCF****,* ***FANCG****,* ***FANCI****,* ***FANCL****,* ***FANCM****, FLCN, GFI1, GPC3, HAX1, HOXB13, KIF1B, MAX, MEN1, MET, MLH1, MRE11A, MSH2, MSH6, MUTYH,* ***NBN****, NOP10, NSD1, NUDT1, OGG1,* ***PALB2****, PMS1, PMS2, POLH, PRSS1, PTEN,* ***RAD50****,* ***RAD51C****,* ***RAD51D****,*

RET, SDHA, SDHAF2, SDHB, SDHC, SDHD, ***SLX4****, SMAD4, STK11, TERT, TMEM127, TP53, TSC1, TSC2, VHL, WAS, WT1, XPA, XPC* and ***XRCC2***.

Pipeline analysis of raw data was performed by Gene Systems (Valencia, Spain), which encompassed: Reads were aligned against the human reference genome version GRCh38/hg38 using BWA aligner. Low quality reads and PCR duplicates were removed from BAM formatted file. Coverage and efficiency of targeted enrichment system were evaluated. Variant calling was performed using a combination of VarScan and GATK, along with in-house scripts to combine and filter variants. Identified variants were annotated using Ensembl database. The candidate deleterious variants for *ATM* gene were confirmed by Sanger sequencing in index cases samples. Twenty other breast and/or ovarian cancer susceptibility genes or candidate susceptibility genes included in the panel were subsequently analyzed in the 392 index cases. The genes are highlighted in bold in the list provided before, only unequivocally deleterious mutations were taken into account for this specific analysis.

4.2.10 Immunohistochemistry

For evidencing loss of ATM protein expression in BC tumor, paraffin embedded tissue, an initial automated dewaxing and rehydration step followed by heat-induced (100C for 20 min) antigen retrieval (15 min, Bond Enzyme Pretreatment Kit, Leica Biossytems) was performed. The slides were subsequently incubated with 3% hydrogen peroxide (5 min), proper primary antibody peroxidase (clone Y170, Abcam, Cambridge, MA, USA) diluted 1:250 (30 min), a post-primary blocking reagent (to prevent nonspecific polymer binding) (8min), horseradish peroxidase (HRP)-labeled rabbit anti-mouse Ig polymer (8 min) and diaminobenzidine susbtrate (10 min). All reagents were components of the Bond Polymer Refine detection system (Leica Biosystems).

4.2.11. Genotyping of 50 Spanish MBC cases

Targeted regions selected upon the analysis of the MBC family through WES were amplified by PCR technology using a suitable primer pair [Supplementary Table 2]. The products were sequenced using the Big Dye Terminator Cycle sequencing kit on an ABI 3730xl Sequencer (Applied Biosystems) to evaluate the presence and status of the variant in a set of 50 MBC cases from Spanish origin.

4.2.12. KBioscience Competitive Allele-Specific PCR genotyping system (KASP) Genotyping technology

For the development of the assays, KASP probes were designed and validated by the company as KOD (KASP On Demand) probes. DNA extraction was performed with Qiagen QIAamp DNA Blood mini kit (cat No 51106) for the samples which were extracted from buffy coat samples and

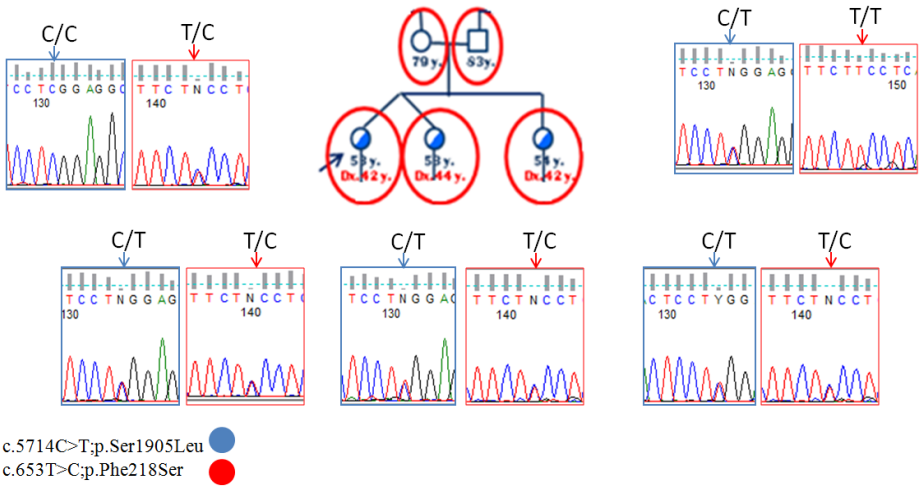
Qiagen QIAamp DNA Blood maxi kit (cat No 51194) for the samples extracted from whole blood. DNA was quantified using PicoGreen ds Quantitation Reagent and its quality checked in a 2% agarose gel. After that, DNA was sampled at a concentration of 10 ng in a final volume of 1 ul, followed by the same volume of Master Mix ROX 2X Buffer and 0.040 ul of KASP KOD probes per samples, dispensed using an Echo Cytometer machine in an ABI PPi plate for 384 samples. Controls with known status of the variants were included to validate the assay and discriminate the allelic groups, as well as adequate non template controls (NTCs). Plates were sealed, properly mixed and spinned down. Standard KASP thermal cycling protocol consists of 10 cycles of touchdown PCR (annealing 61-55C, decreasing 0.6C per cycle), then 26 cycles of standard 2-step PCR at the lower annealing temperature (55C). The other variation of program was set at a touchdown of 68C-62C. The difference of the protocol used resides in the content of G/C and the size of the region of interest to amplify. For the lecture, three additional cycles were used to refine the clusters in the allelic discrimination plot. Genotyping lectures were obtained in a QuantStudio 6 Flex System Machine. Data analysis was performed using the Quant Studio Real-Time PCR Software version 1.4.

5. Results

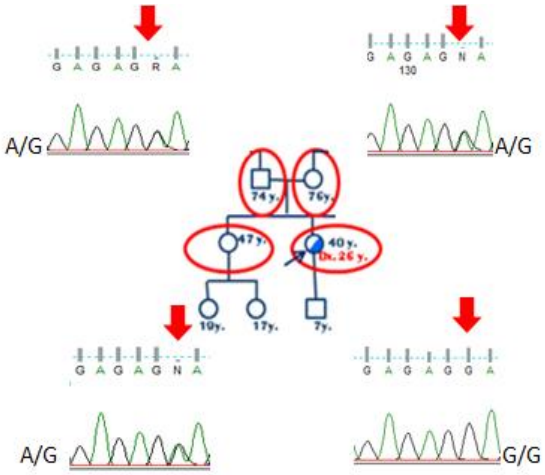
5.1. Exploration of the recessive model of inheritance in BRCAX female BC families

Our initial hypothesis for the exploration of the families was the recessive model of inheritance. In order to detect potential susceptibility genes in hereditary breast cancer families (BRCAX) with this model of inheritance, we applied two approaches: 1) look for variants in homozygous state in BC affected siblings in the families, each variant coming from each parent and 2) look for variants as compound heterozygous (different variants that affect the same gene) in BC affected siblings and each variant coming from each parent. Specifically, in family 1, we looked for variants in homozygous and compound heterozygous state present in the three BC affected members [Figure 9A]. In family 2, we looked for presence of variants in homozygous [Figure 9B] or compound heterozygous state [Figure 9C] in BC affected sister, heterozygous state or homozygous for the reference in the healthy sister (this member of the family was assumed as healthy since she had not developed BC at 47 years of age) and heterozygous state in the parents. In family 3, due to technical constraints we did not have the exome from the parents, although we looked for variants in homozygous and compound heterozygous in BC affected patient from whom we had the exome. Nonetheless, after the application of our priorization filters (described in Methods), we did not find any variant that adjusted to the analyzing criteria applied. As a consequence, no variants in a recessive model of inheritance were found for this family. Family 4 (a hereditary male BC family) was analyzed separately and we will refer to it in a separate section. Of importance, exploration of genes related with roles in DNA repair, tumoral suppression, cell growth, cell signaling, cell survival; as well as genes implicated in hormone metabolism and signaling pathways were prioritized in order to increase the possibility to find a novel susceptibility BC gene. The list of candidate variants for family 1 and 2 are shown in Table 3, pathways and main cellular functions are shown in Supplementary Table 1.

A)



B)



C)

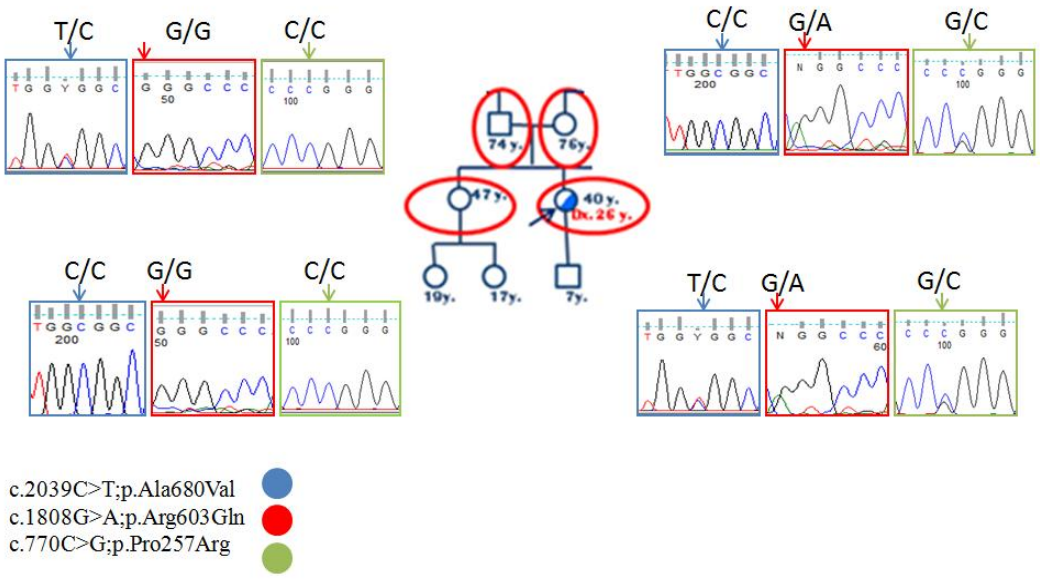


Figure 9. Example of the segregation analysis showing the recessive model of inheritance in family 1 and 2. A) Compound heterozygous model of inheritance showing two variants in *DOCK3* gene present in the three affected sister with BC, each variant inherited from one of the parents. B) *AKR1C3* (c.230A>G; p.Glu77Gly) is the only gene for which we found a variant in homozygosis in the affected sister and in heterozygosis the parents and healthy sister. C) Compound heterozygous model of inheritance showing three variants in *FBF1* gene present in the affected sister with BC and absent in the healthy sister. Two are inherited from the mother and one from the father. Affected members with BC are represented with half colored circles; Dx, Age of diagnosis; arrow, index case; red circle, analyzed by WES.

Table 3.

Candidate variants found under the recessive model of inheritance by WES in BRCAX families

Fam	Gene	Chromo-somal position	Functional class	Ref/Alt	rs	GMAF[1]	EXAC[2]	Variant
1	*DOCK3*	51418611	Missense	C/T	rs201507848	0.0012	0,0003721	S1905L
		51127722	Missense	T/C	NR	NR	NR	F218S
2	***AKR1C3***	**5138747**	**Missense**	**A/G**	**rs11551177**	**0,03672**	**0,05018**	**E77G**
	FBF1	73916166	Missense	C/T	NR	NR	0.00034*	R603Q
		73915803	Missense	G/A	rs113062332	0.0074	0,02305	A680V
		73922167	Missense	G/C	rs201197761	NR	0,0002256	P257R
	TOPAZ1	44285446	Missense	A/G	rs17076541	0,0058	0.003706*	Q483R
		44286015	Missense	C/G	rs17646517	0.0080	0.01783*	P673A
		44286384	Missense	A/G	rs17076545	0.0110	0.005698*	K796E
		44373498	Missense	C/G	rs533942526	0.0002	0.0000922*	H1692D
	GLE1	131271171	Missense	G/A	rs138871311	0.0002	0.0001977	C39Y
		131271281	Missense	G/A	NR	NR	0.0000082	A76T
	KLB	39448672	Frameshift	CTCTC/C	NR	NR	0.002272	F777del
		39450295	Missense	G/A	rs143809363	0.0006	0.003605	V1042I

Bold marks, unique variant found in homozygous state according to recessive model of inheritance. The rest of the variants were found as compound heterozygous. NR, Not reported. [1]GMAF based in 1000Genomes reports. [2]Total include European (non Finnish and Finnish), Latino, South Asian, African and East-Asian populations. .*For this variant, a note in EXAC is given: This site is covered in fewer than 80% of the individuals in ExAC, which may indicate a low-quality site

5.2. Exploration of genes related with DNA maintenance and repairing systems

Many genes that confer high-/moderate- susceptibility for BC are related with DNA maintenance and DNA repairing systems (Mavaddat et al., 2010). This is why, regardless of the model of inheritance, further priority was given to genes with a well-established role in DNA repairing pathways. We looked for the presence of variants in homozygous or heterozygous status affecting known genes that are related with these pathways in family BC affected members from family 1, 2 and 3. For this list of genes, information from *in silico* predictors was not taken into account; as well, we considered synonymous variants and checked them all. The only filter that we considered was the frequency of the variants in population (see methods).

The list of genes explored was: *UNG, SMUG1, MBD4, TDG, OGG1, MUTYH, NTHL1, MPG, NEIL1, NEIL2, NEIL3, APEX1, APEX2, LIG3, XRCC1, PNKP, APLF, PARP1, PARP2, PARP3, MGMT, ALKBH2, ALKBH3, TDP1, TDP2, MSH2, MSH3, MSH6, MLH1, PMS2, MSH4, MSH5, MLH3, PMS1, PMS2L3, XPC, RAD23B, CETN2, RAD23A, XPA, DDB1, DDB2, RPA1, RPA2, RPA3, ERCC3, ERCC2, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, CDK7, CCNH, MNAT1, ERCC5, ERCC1, ERCC4, LIG1, ERCC8, ERCC6, UVSSA, XAB2, MMS19, RAD51,*

RAD51B, RAD51D, DMC1, XRCC2, XRCC3, RAD52, RAD54L, RAD54B, BRCA1, SHFM1, RAD50, MRE11A, NBN, RBBP8, MUS81, EME1, EME2, GIYD1, GIYD2, GEN1, FANCA, FANCB, FANCC, BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, BRIP1, FANCL, FANCM, PALB2, RAD51C, BTBD12, FAAP20, FAAP24, XRCC6, XRCC5, PRKDC, LIG4, XRCC4, DCLRE1C, NHEJ1, NUDT1, DUT, RRM2B, POLB, POLG, POLD1, POLE, PCNA, REV3L, MAD2L2, REV1L, POLH, POLI, POLQ, POLK, POLL, POLM, POLN, FEN1, FAN1, TREX1, TREX2, EXO1, APTX, SPO11, ENDOV, UBE2A, UBE2B, RAD18, SHPRH, HLTF, RNF168, SPRTN, RNF8, RNF4, UBE2V2, UBE2N, H2AFX, CHAF1A, SETMAR, BLM, WRN, RECQL4, ATM, TTDN1, DCLRE1A, DCLRE1B, RPA4, PRPF19, RECQL, RECQL5, HELQ, RDM1, OBFC2B, ATR, ATRIP, MDC1, RAD1, RAD9A, HUS1, RAD17, CHEK1, CHEK2, TP53, TP53BP1, RIF1, TOPBP1, CLK2,l PER1, as reported in https://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html#BER, last accessed December 2014. The results of this exploration are shown in Table 4

Table 4. Candidate variants found in DNA repairing genes found by WES in BRCAX families

Family	Gene	Chromo-somal position	Functional class	Ref/Alt	rs	GMAF[1]	EXAC[2]	Variant
1	*POLQ*	121208275	Missense	A/G	NR	NR	NR	I1168T
	CHEK2	29083961	Missense	C/A	NR	NR	0,0002315	R562L
	POLG	89865073	Missense	T/C	rs41549716	0,0022	0,006277	Y831C
2	*POLK*	74892973	Missense	A/C	rs185752953	0.0006	0.001024	N819H
	RECQL5	73623704	Missense	G/C	rs200560792	NR	0,0001392	S958R
	RECQL5	73626708	Missense	G/A	NR	NR	0.00002491	T570I
3	*ATM*	108173695	Frameshift	c.5441del	NR	NR	NR	L1814fsX14

NR, Not report. [1]GMAF based in 1000Genomes reports. [2]Total include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations.

In family 1, we found a variant that affected a well-established moderate HBOC gene, *CHEK2*.

We found variants affecting DNA polymerases theta (family 1) and kappa (family 2), which are mediators of microhomology-mediated end-joining (MMEJ), an alternative non-homologous end-joining (NHEJ) machinery triggered in response to double-strand breaks and catalyzes translesion DNA synthesis, which allows DNA replication in the presence of DNA lesions, respectively (Ceccaldi et al., 2015; Fischhaber et al., 2002)

We identified two germline variants affecting gene *RECQL5* in family 2, for which relation with HBOC has not been explored, although other members of the family of genes have already been described as potential HBOC susceptibility genes (Cybulski et al., 2015; Sun et al., 2015). Both variants were inherited from the mother. The variants were present in the affected sister with BC, but not in the healthy sister. Variants were missense: c. 1709C>T, p. T570I and c.2874C>G, p.S958R and would show incomplete penetrance, due to the presence in the mother (I.2) who is not affected with BC. Segregation analysis revealed the absence of both variants in the father (I.1) and in the healthy sister (II.1) [Figure 10a, 10b].

c.1709C>T, p.T570I was not reported in LOVD database, meanwhile the variant c.2874C>G, p.S958R was reported from a panel of colon cancer genes with no phenotype associated in the heterozygous carrier (genre not specified) (Genomic Variant #0000148990, http://databases.lovd.nl/). None of these variants is reported in COSMIC. All data was last accessed March 2017. The variant c.2874C>G, p.S958R is localized in the end of the exon 19 and was predicted to affect splicing based on the predictions given by the splicing module integrated in Alamut® Visual 2.7. Through analysis of the cDNA of healthy and affected sister (II.1 and II.2), we realized that the variant leads to aberrant splicing, causing partial skipping of exon 19 [Figure 10c] which is predicted to generate a truncated product. This event was not present in the healthy sister (II.1).

Theoretically the variant would provoke the loss of PIM (PCNA-Interacting Motif, residues 964-971) and half of Set2-Rpb1-interacting (SRI) domain (residues 900-991) which is needed for interaction with Ser2-Ser5-phosphorylated CTD (Carboxi Terminal Domain) of PolII required for the elongation state of transcription (Kanagaraj et al., 2010; Kassube et al., 2013). As this was a potentially deleterious mutation and because of all the pathways in which RECQL5 is implicated, we explored its role as a novel BC susceptibility allele.

In family 3, we identified a variant that affected *ATM*, a known moderate susceptibility allele for BC. Of importance, the variant is a frameshift mutation (c.5441del, L1814fsX14) that leads to the premature truncation of the protein. As this mutation is clearly deleterious, it was considered as the causal variant of the disease in family 3. Segregation analysis revealed that the mutation was present in index case´s sister affected also with BC at 35 years of age, nonetheless the variant was not fully penetrant since the mother (I.1) carried the variant without any cancer diagnosis at 77 years of age [Figure 11a, 11b].

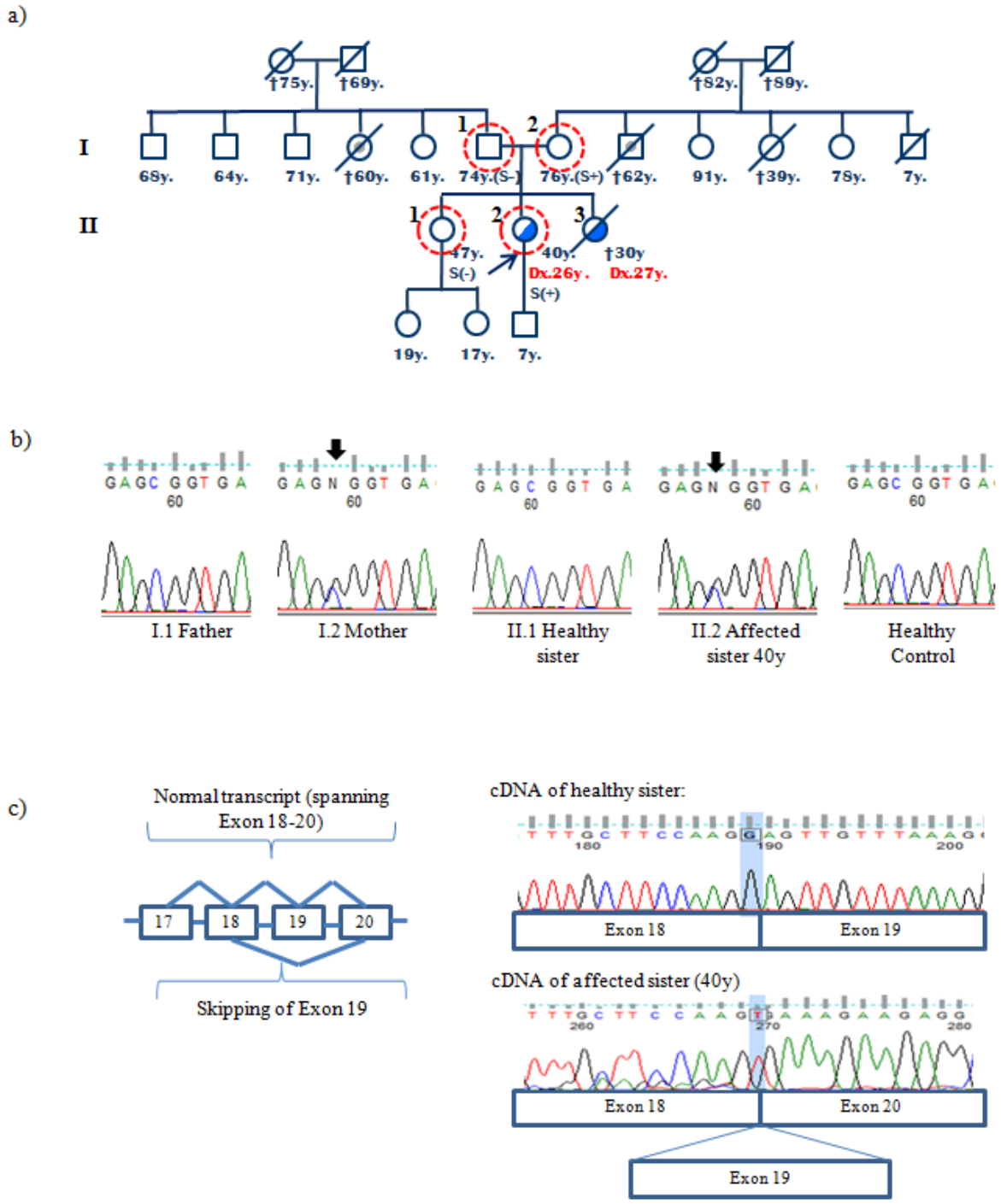


Figure 10. a) Family 2. Affected members with BC are represented with half colored circles. Dx, Age of diagnosis; arrow, index case; red dashed circle, analyzed by WES; S, result of segregation analysis; LOH, result of loss of heterozigosity; Other types of cancer are represented with grey dots inside circles. b) Family segregation analysis of *RECQL5* variant (c.2874C>G; p. S958R), the other variant (c.1709C>T; p.T570I) has the same pattern (data not shown), performed by Sanger sequencing in indicated members of the family. c) Representation of aberrant splicing in II.2. Sequences of normal and aberrant splicing with partial skipping of exon 19 seen in cDNA from healthy and affected siblings is presented.

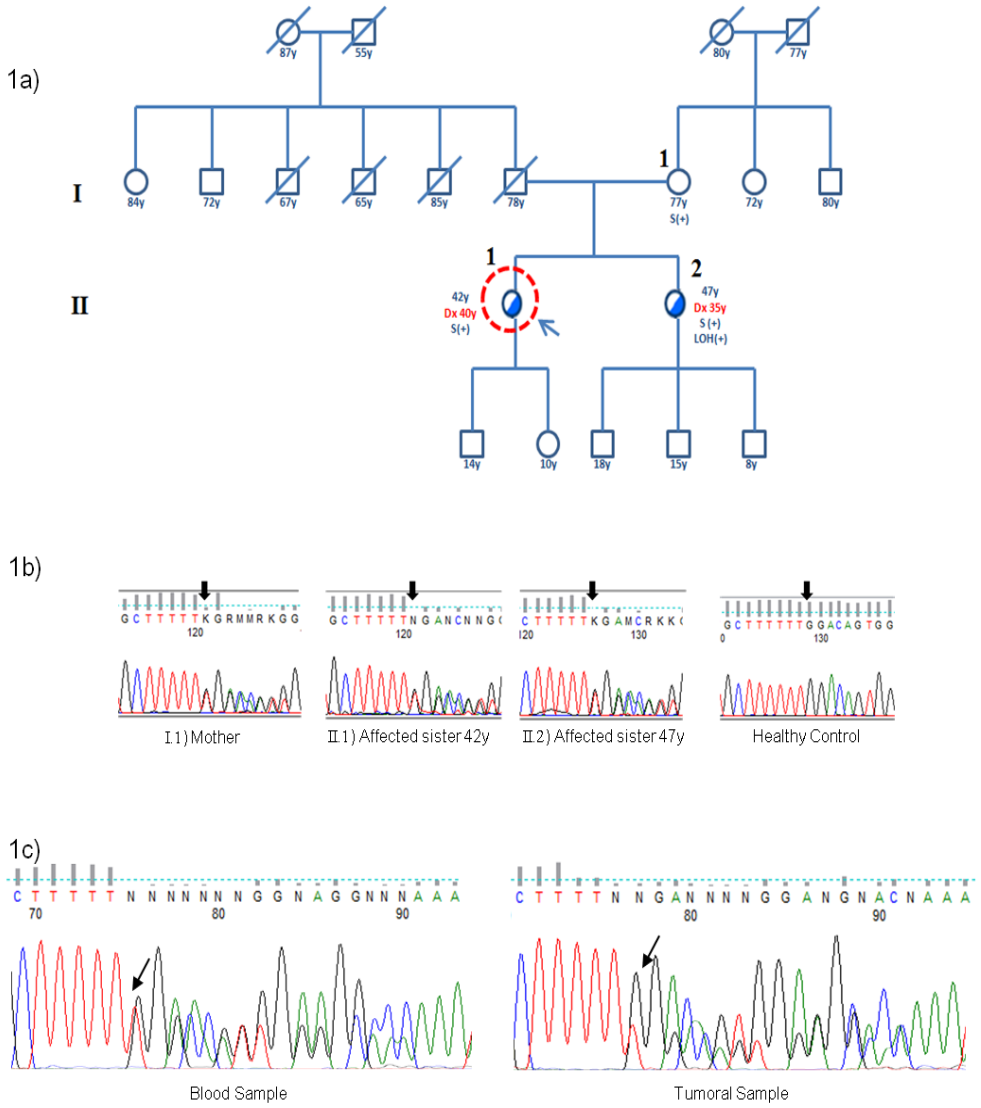


Figure 11. a) Family 3. Affected members with BC are represented with half colored circles. Dx, Age of diagnosis; arrow, index case; dashed circle, analyzed by WES; S, result of segregation analysis; LOH, loss of heterozigosity. b) Family segregation analysis of *ATM* truncating mutation (c.5441delT; p. p.Leu1814Trpfs*14) performed by Sanger sequencing in indicated members of the family. Healthy control does not belong to the family and has none familial story of HBOC. c) Loss of heterozigosity. (Left) Blood sample from affected sister at 35y (II.2). (Right) Tumoral tissue from affected sister at 35y, loss of normal allele as depicted by prominent peak of altered allele.

According with the pathogenicity of the mutation, tumoral DNA from paraffin-embedded br**e**ast tumor from one of the affected sisters (II.2) demonstrated loss of wild type allele [Figure 11c]. In addition, IHQ analysis of the same tumor revealed very weak or absent expression of ATM in the tumoral cells compared with adjacent normal or immune cells, supporting the causality of the mutation, associated with a dominant model of inheritance [Figure 12]. Of interest, the mutation had not been previously reported as a germline pathogenic variant, however, it had been found somatically, in combination with a *PTEN* somatic mutation, in an advanced stage ovarian cancer (Pennington et al., 2014).

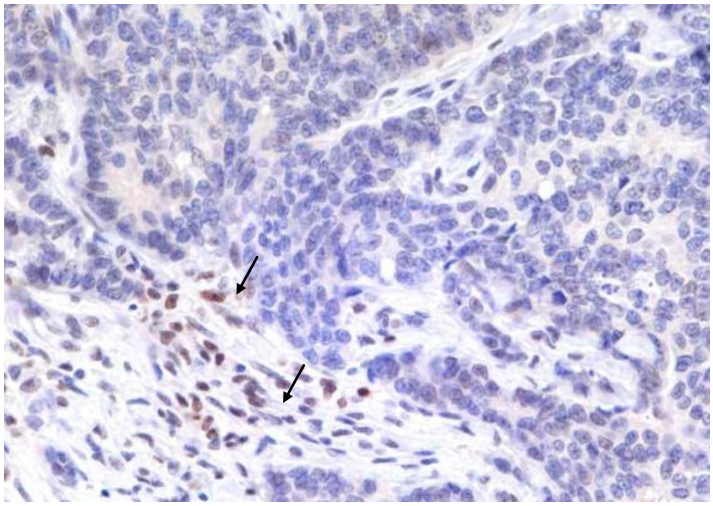


Figure 12. IHQ analysis of the ATM protein in a section of the breast tumor of individual (II.2). Nuclear ATM expression is absent in tumoral cells (Marked with black arrow) in comparison with normal counterparts.

Noteworthy, all the other variants found while exploring genes related with DNA maintenance and repairing systems were heterozygous. These findings, together with the fact that a clearly deleterious variant in *ATM* explained the disease in BRCAx family 3, pointed to the fact that we could not rule out other models of inheritance because, at least in one family, we were facing a dominant model of inheritance. As a consequence, we explored this model of inheritance and the results are presented in the next section.

5.3. Exploration of the dominant model of inheritance in BRCAX families

While searching for variants located in genes related with DNA maintenance and repairing systems, we found seven variants which would fit a dominant model of inheritance; the same model of inheritance that is followed by the two major BC high susceptibility genes, *BRCA1* and *BRCA2*. Given this result, we wanted to complete the panorama by exploring the rest of the genes

(not related with DNA repairing systems) in a dominant model of inheritance (heterozygous variants), in order to find novel variants and/or genes that could increase susceptibility to HBOC [Table 5].

Table 5. Candidate alleles found under the dominant model of inheritance by WES in BRCAX families

Fam	Gene	Chromo-somal position	Functional class	Ref/Alt	rs	EXAC[1]	Variant
1	*KANK1*	742215	Frameshift	g.742217delG	NR	0.0000083	A1237fsX11
	CDH15	89257818	Frameshift	g.89257821delC	NR	NR	P439fsX38
	ACTR8	53902781	Missense	G/T	rs201450445	0.000033	R614S
	CERS5	50561013	Missense	A/T	NR	NR	L23Q
	KAT5	65480436	Missense	C/G	NR	NR	D97E
	MTUS1	17504536	Missense	T/G	NR	NR	N1185T
	PLEC	144991514	Missense	G/A	rs201069314	0.00047	R4296C
	DDX11	31238059	Missense	A/G	rs2536756	0.002273	R213G
	NISCH	52521676	Missense	G/A	NR	0.0000165	R723H
2	*DEPDC1B*	59938657	Missense	T/C	rs199716657	0.000046	M247V
	SPRED2	65561832	Missense	C/T	NR	0.0000082	D94N
	GPRC5B	19883558	Missense	G/T	NR	0.000016	L204I
	SYNE1	152476127	Missense	C/T	NR	0.0000165	R8010H
	HBP1	106836378	Missense	A/C	NR	0.000058	Q389H
	AHNAK	62293925	Missense	C/A	rs149615783	0.00074	G2655V
	KLF15	126070801	Missense	T/C	NR	0.000036*	K322R
	MYBBP1A	4455198	Missense	C/T	rs138633396	0.0011	G334R
	HHAT	210761290	Frameshift	ACT/A	NR	NR	L366fsX86
	RANBP10	67840391	Missense	C/T	NR	NR	D17N
	CAPN2	223900358	Missense	G/A	rs375899944	0.00011*	A6T
	CLSPN	36228012	Missense	G/A	rs141350492	0.00016	T272M
	PTPRG	62189151	Missense	C/A	rs201820508	0.00043*	S561Y

NR, No Report. [1]Total include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations. These list of genes do not include genes that are related with DNA repairing systems, as this has been already explored in the previous section.*For this variant, a note in EXAC is given: This site is covered in fewer than 80% of the individuals in ExAC, which may indicate a low-quality site

As with the recessive model of inheritance, in the dominant model of inheritance, exploration of genes related with roles involved in tumoral suppression, cell growth, cell signaling, cell survival; as well as genes implicated in hormone metabolism and signaling pathways were prioritized in order to increase the possibility to find a novel susceptibility BC gene. The pathways and main cellular functions where the genes are implicated together with the model of inheritance studied are shown in Supplementary Table 1.

5.4. Case-control association study of putative susceptibility variants associated with Hereditary Breast and Ovarian Cancer in Spanish population

Case-control association study was performed to establish the frequency of the variants in cases and in controls, so that variants that were present with higher frequency in controls rather than in cases were ruled out as this would be associated with a protective effect. This served us to narrow the number of candidate variants and select the most interesting ones with a putative association with HBOC in BRCAX families from Spanish population.

The initial number of candidate variants selected through WES data mining of family 1, 2 and 3, taking into account the recessive model, the DNA repairing genes and the dominant model of inheritance was 43. Some of them were discarded through case-control association due to a frequency that did not fit with the model of inheritance studied and/or higher recurrence in controls against cases. These included the following: *POLG* (c.2492A>G; p.Y831C), *FBF1* (c.2039C>T; p.A680V), *MYBBP1A* (c.1000G>A; p.G334R), *POLK* (c.2455A>C; p.N819H), *NISCH* (c.2168G>A; p.R723H), *CLSPN* (c.815C>T; p.T272M) and *TOPAZ1* (c.2017C>G; p.P673A). For other variants, the probes from the OpenArray did not showed results, so other types of approaches were applied in order to get the case-control association information. Such was the case with variants located in *AKR1C3* (c.230A>G; p.E77G) and *CHEK2* (c.1685G>T, p. R562L). For the former, even the Taqman assay did not showed results but for the latter, it was finally discarded due to its high frequency that do not fit with the dominant model of inheritance that was studied for this variant. Some other variants were discarded since the first exploration, because the probes designed could not efficiently discriminate among allelic clusters due to the position of the variants, as is the case for *SPRED2* (c.280G>A; p.D94N) and *DDX11* (c.637A>G; R213G). Due to technical limitations, some variants could not be included in the initial OpenArray study and will be included in further explorations as is the case for the variants located in *KLB* (c.2329_2331del; p. F777del; c.3124G>A; p. V1402I) and *TOPAZ1* (c.5074C>G; p.H1692D). Finally, some others were eliminated since the very beginning, because the design of the probe did not worked either in the CEGEN facility or in the company where it was designed as for *CAPN2* (c.16G>A; p.A6T) and *PTPRG* (c.1682C>T; p.S561Y) The summary of the variants that were finally not included in larger cohorts can be found in Supplementary Table 4.

In table 6, the number of heterocigotes found for the 25 candidate variants in 1500 BRCAX cases and 500 controls is shown. Variants absent in cases and controls or those more prevalent in cases than controls were selected as candidates for further analysis. Of notice, the p-values were not valuable for setting selection thresholds as the variants are quite rare in the population. In a

similar way, it was not possible to calculate p-value in variants where no heterozygotes were found both in cases and controls.

Table 6. Final candidate variants selected through case-control association study in Spanish population

Gene	Chr	Variant	rs	P value	Cases	Total Cases	Controls	Total Controls
KANK1	9	c.3709del	NR	0,495	0	1484	1	604
CDH15	16	c.1316del	NR	NA	0	1491	0	600
ACTR8	3	c.1840C>A	rs201450445	NA	0	931	0	502
CERS5	12	c.68T>A	NR	NA	0	1495	0	597
KAT5	11	c.291C>G	NR	NA	0	1498	0	606
MTUS1	8	c.3554A>C	NR	NA	0	1461	0	600
PLEC	8	c.12886C>T	rs201069314	0,256	12	1493	2	604
FBF1*	17	c.1808G>A	NR	0,33	5	1464	0	582
		c.770C>G	rs201197761	0,584	4	1491	0	601
RECQL5	17	c.2874C>G	rs200560792	NA	0	1491	0	601
		c.1709C>T	NR	NA	0	1491	0	588
DEPDC1B	5	c.739A>G	rs199716657	1	1	1427	0	569
GPRC5B	16	c.610C>A	NR	1	2	1488	0	597
SYNE1	6	c.24029G>A	NR	NA	0	1494	0	651
HBP1	7	c.1167A>C	NR	0,575	11	1474	3	603
AHNAK	11	c.7964G>T	rs149615783	0,443	6	1176	1	543
KLF15	3	c.965A>G	NR	NA	0	1490	0	603
HHAT	1	c.1096_1097del	NR	NA	0	1475	0	589
RANBP10	16	c.49G>A	NR	NA	0	1484	0	591
POLQ	3	c.3503T>C	NR	1	1	1495	0	595
GLE1	9	c.116G>A	rs138871311	0,583	4	1478	0	590
		c.226G>A	NR	NA	0	1494	0	602
TOPAZ1*	3	c.1448A>G	rs17076541	0,527	10	1491	2	602
		c.2386A>G	rs17076545	0,177	15	1472	2	585
ATM	11	c.5441delT	NR	NA	0	1477	0	589

Bold marks, variants with phenotypes described associated with cases. NR, not reported, NA, not applicable; Underlined, refers to the variants initially studied under the recessive model of inheritance, the rest belong to the exploration of the DNA repairing and maintenance related genes and the dominant model of inheritance. *Variants that remained in case-control association study in *TOPAZ1* and *FBF1*, were inherited from one parent (from maternal branch), the other variants (2 for *TOPAZ1* and one for *FBF1*) were discarded because of technical constraints or a frequency that does not fit the model studied. Hence, these variants would be studied in the dominant model of inheritance. Though, priority for further study would be set in the genes with higher impact effect and/or function related with cellular signaling pathways, DNA repair and hormone metabolism. NR, no report.

In table 7, we show the phenotypes associated with the variants which were reported in one or more cases, these are marked in the previous table in bold mark. It is interesting to emphasize some examples: variants in *PLEC* c.12886C>T and *HBP1* c.1167A>C, where >70% are breast cancer only cases in BRCAX families.

Table 7. Phenotypes associated with BRCAX index cases carrying candidate BC susceptibility alleles

Gene	Variant	rs	Cases	Phenotype
PLEC	c.12886C>T	rs201069314	12	8BC, 1MBC, 1OC, 2HBOC
FBF1	c.1808G>A	NR	5	2 BC, 2OC, 1HBOC
FBF1	c.770C>G	rs201197761	4	2BC, 2OC
DEPDC1B	c.739A>G	rs199716657	1	1BC
GPRC5B	c.610C>A	NR	2	2 BC
HBP1	c.1167A>C	NR	11	9BC, 1HBOC, 1BiBC
AHNAK	c.7964G>T	rs149615783	6	4BC, 1BiBC, 1OC
POLQ	c.3503T>C	NR	1	1BiBC
GLE1	c.116G>A	rs138871311	4	3 BC, 1 HBOC
TOPAZ1	c.1448A>G	rs17076541	10	7BC, 1OC, 2HBOC
TOPAZ1	c.2386A>G	rs17076545	15	11BC, 1OC, 2HBOC, 1HBiBOC

NR, no report. BC (Breast cancer); BiBC (Bilateral breast cancer); OC(Ovarian Cancer); MBC(Male BC)

In summary, in this first part of the thesis, we have studied three BRCAx families presenting an apparent recessive model of inheritance. After the exploration of the recessive model of inheritance, the DNA repairing genes and the dominant model of inheritance, we have found that two families are most probably explained by the dominant model of inheritance. Although in family 1, we still do not have a conclusive variant, we have a list of variants for further study as potential BC susceptibility alleles. In family 2 we found novel mutations in *RECQL5* (c.1709C>T, p. T570I and c.2874C>G, p.S958R) as a potential novel BC susceptibility gene, as well as a list of other candidate variants associated with the dominant model to further study as BC susceptibility which were not ruled out because the genes are implicated in quite interesting pathways and/or its cellular function points toward a potential tumor suppressor, which require in-depth studies. In family 3 the BC causality was proven associated with a mutation in *ATM* (c.5441delT; p.L1814WfsX14). Also, from the case–control association study performed in 1500 BRCAx cases and 500 controls of Spanish population, we have 25 variants selected to include them in higher number of cohorts to establish significant association between BC risk and the presence of the variants.

The findings in family 2 and 3, lead to the second part of the thesis, where we performed full coding sequencing and exon-boundaries analysis of *RECQL5* and *ATM*, in order to establish the frequency and spectrum of mutation of both genes in Spanish population.

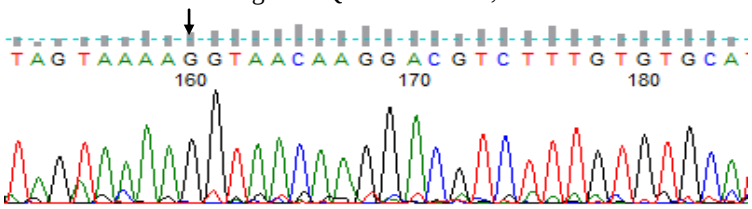
5.5. Massive sequencing of *RECQL5* gene in 700 BRCAX breast cancer only-cases and 754 controls from Spanish population

Due to our findings in family 2 with two variants affecting BC index case in *RECQL5* and considering the seminal role of this gene in several cellular pathways: DNA repairing systems and maintenance, DNA replication, transcription and repairing (Paliwal et al., 2014; Popuri et al., 2013; Kanagaraj et al., 2010), genome stability (Islam et al., 2012), telomere function, homologous recombination (Pellatt et al., 2013; Hu et al., 2007), we decided to perform full coding and exon boundaries sequencing of the gene in a cohort of 700 BRCAX BC only-cases and 754 controls in Spanish population, in order explore its role in BC.

The full coding sequencing of *RECQL5* revealed the presence of three germline pathogenic variants, present in unrelated BC-only BRCAX cases (Figure 14). Two of them are clearly deleterious mutations, as they are frameshift variants that lead to a truncated product (c.657delC, p. C220fsX15; c.2393dupC, p, M799fsX24), whereas the third one is a missense mutation (c.130G>A, p.G44S) that is located in the last nucleotide of exon 2 and predicted to affect splicing. Access to the DNA from other family members from BC index case harboring *RECQL5* c.130G>A, G44S was possible. Segregation analysis from dizygotic twin sister and another sister, affected with BC at 54 and 53 years old, revealed absence of the variant. Due to the incomplete segregation of the variant in other members of the family with BC, we hypothesize that this could be associated with a moderate risk; as well, other gene that confers susceptibility could be related with the disease and germline variants in RECQL5 could be modifying the risk for BC. We have to take into account that is was not possible to gather sample from the mother, which would be interesting in light of clarifying the role of RECQL5 missense variant in this family, whom could be also harboring the variant.

At cDNA level, in the BC index case, we explored if there was aberrant splicing. Surprisingly, the missense variant c.130G>A, G44S was not present in heterozygosity at cDNA level [Figure 13] although we were not able to detect which was exactly the effect of exon skipping or inclusion. Other SNPS along exon 9-10 (c.1439A>G) and exon 16 (c.2217C>T), revealed that at these positions, heterozygosity of selected SNPs was present at cDNA level. Although not fully characterized, we can assume that the variant affects severely the mRNA splicing process.

a) cDNA from BC index case harboring RECQL5 c.130G>A, G44S



b) DNA from BC index case harboring RECQL5 c.130G>A, G44S

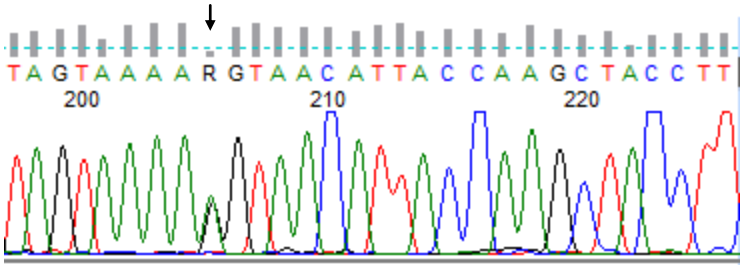


Figure 13. cDNA and DNA from BC index patient harboring RECQL5 c.130G>A, G44S, loss of heterozigosity is seen at cDNA level [a] in contrast with DNA [b].

We could not extend segregation analysis to other members of the families of BC indexes case, harboring (c.657delC and c.2393dupC), because access to the samples was not possible. All the germline clearly pathogenic variants were found in heterozygous state in index cases. The phenotype, age of diagnosis and familial antecedents are presented in Table 8. As well, the pedigrees are presented in Figure 14. In all the families there were cases of other cancers different than BC and OC in previous generations. In this regard, family with BC index case carrying *RECQL5* c.2393dupC, M799fsX24, needs further exploration due to the early decease of the mother of BC-affected sibling and the difference on age between first and last daughter, this suggest that some of the siblings could be sons or daughters from a second marriage of the father, which is of importance in segregation analysis of other members of the family, particularly, to detect mutation carriers within the sisters of the BC index case. No clearly pathogenic germline variant was found in the 754 healthy controls explored.

The prevalence of pathogenic germline variants in *RECQL5* gene associated with in BRCAX families with BC-only cases was 0.42% (3/700).

Table 8. *RECQL5* germline deleterious variants and phenotype associated found in 700 Spanish BRCAX BC-only families by NGS

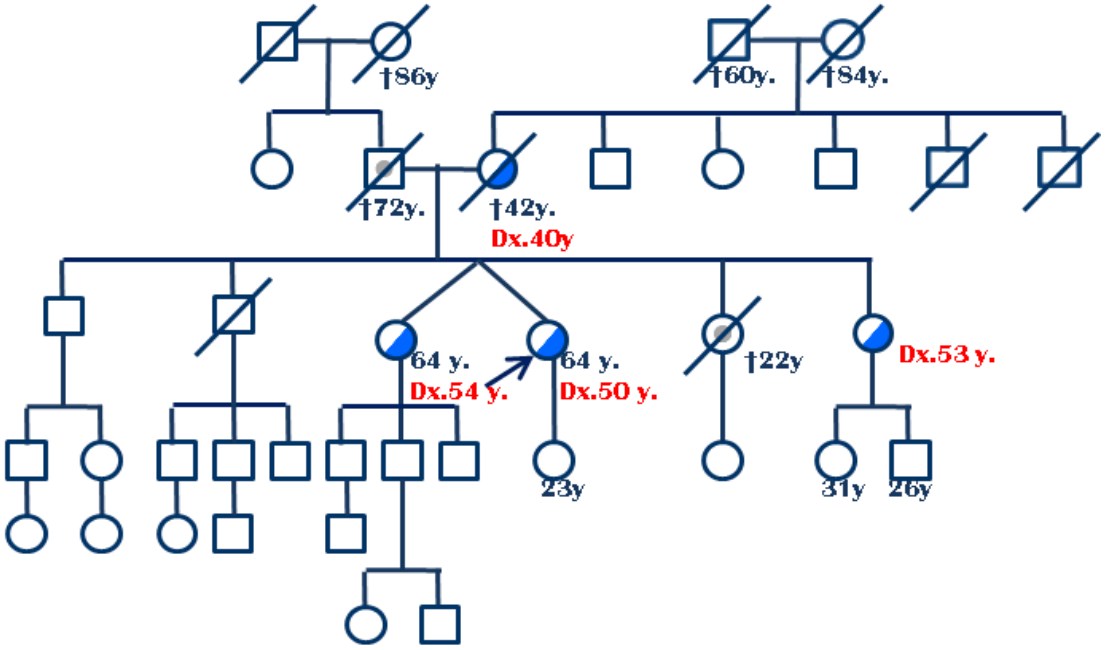
Nucleotide Change[1]	Protein change	Gender	Index Case Phenotype and Age of diagnosis	Other cancers in index case	ClinVar	EXAC[3]
c.2874C>G[2]	S958R	F	BC, 26y	No	NR	0.0001392
c.130G>A	G44S	F	BC, 50y	No	NR	NR
c.657delC	C220fsX15	F	BiBC, 34y, 46y	No	NR	NR
c.2393dupC	M799fsX24	F	3N BiBC, 37y, 39y	No	NR	NR

[1] GenBank reference sequence NM_004259.6 with numbering starting at the A of the first ATG, following the HGVS guidelines, www.hgvs.org/mutnomen. [2]Mutation found by WES was not taken in account to establish population percentages. F (Female), BC (Breast Cancer), OC (Ovarian Cancer), 3NBiBC (Triple Negative Bilateral BC), NR (No Reported). [3] Total values for EXAC include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations.

Rare germline variants in *RECQL5* detected in Spanish population, predicted to have a pathogenic effect (according to the combined information provided by 8 *in silico* predictors and PPS score) are listed in Table 9. We found 4 potentially deleterious *RECQL5* germline variants in cases. The pedigrees of the families are presented in Figure 15, all except one proceeded from external hospitals and we had no access to more information about this family. In controls, we found 5 rare *RECQL5* germline variants.

The location of all the predicted and clearly deleterious mutations in cases and controls are depicted along RECQL5 protein, together with all the site-directed mutagenesis *in vitro* assays that have been performed in RECQL5 to establish transcendence of missense mutations in different domains of the protein [Figure 16C].

A) Family 2099 CNIO with BC index case, harboring *RECQL5* c.130G>A, G44S



B) Family 1481, index case 5431 from San Carlos Clinic Hospital with BC index case, *RECQL5* c.2393dupC, M799fsX24

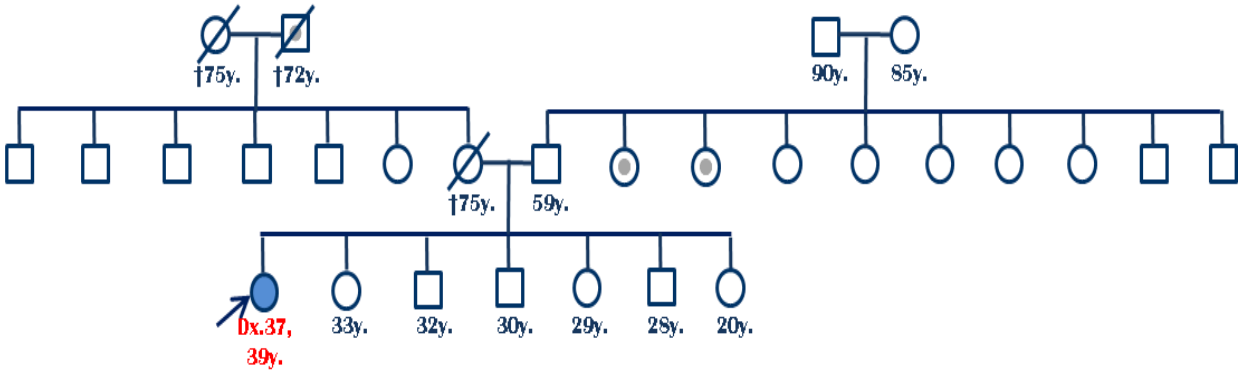


Figure 14. Families harboring *RECQL5* germline deleterious variants. a) Family with BC index case, harboring c.130G>A, p.Gly44Ser. b) Family with BC index case, harboring c.2393dupC, M799fsX24. Affected members with BC are represented with half colored circles and bilateral BC with complete colored circle; Dx, Age of diagnosis; arrow, index case; other types of cancer are represented with grey dots inside circles. When information was available, the type of cancer is specified in the pedigree.

C) Family 988, index case 3337 from San Carlos Clinic Hospital with BC index case, *RECQL5* c.657delC, C220fsX15

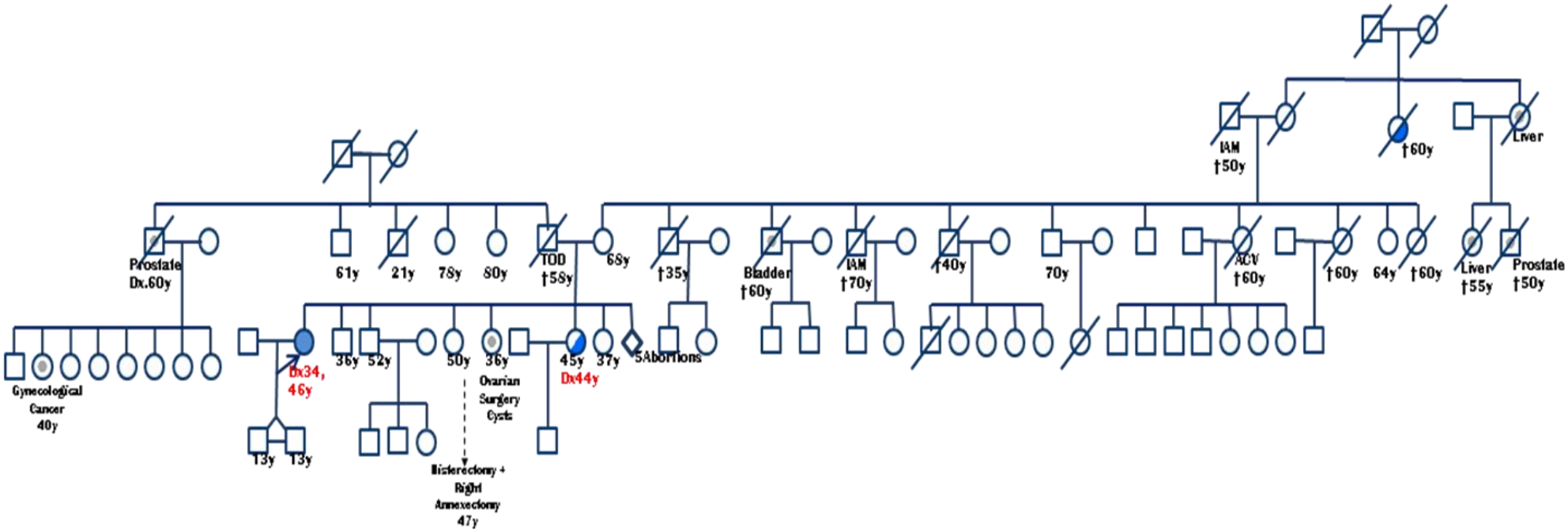


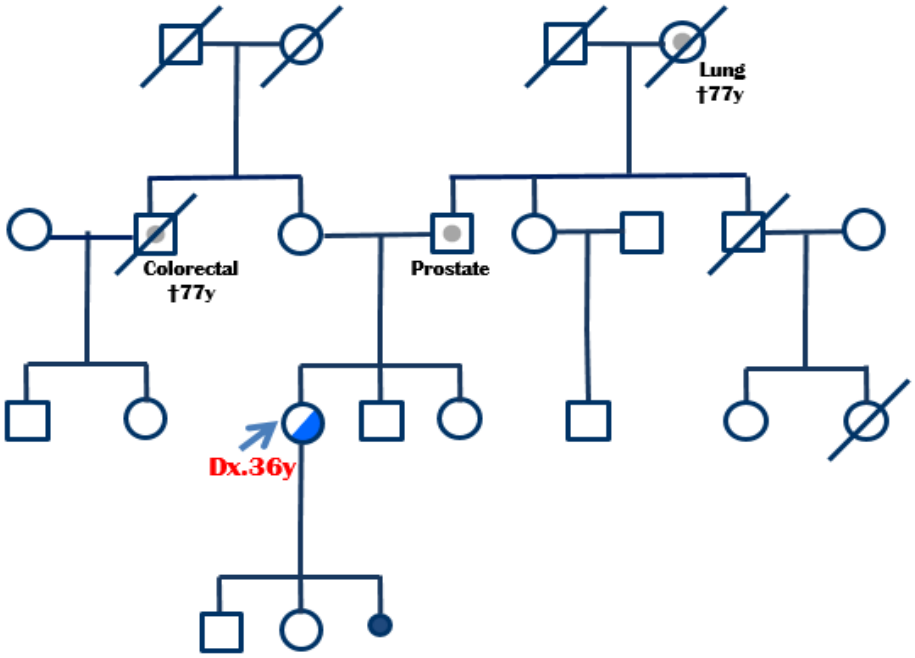
Figure 14. Families harboring *RECQL5* germline deleterious variants. C)Family with BC index case, harboring c.657delC, C220fsX15. Affected members with BC are represented with half colored circles and bilateral BC with complete colored circle; Dx, Age of diagnosis; arrow, index case; other types of cancer are represented with grey dots inside circles. When information was available, the type of cancer is specified in the pedigree.

Table 9. Putative pathogenic *RECQL5* germline variants detected through mutational screening in Spanish population

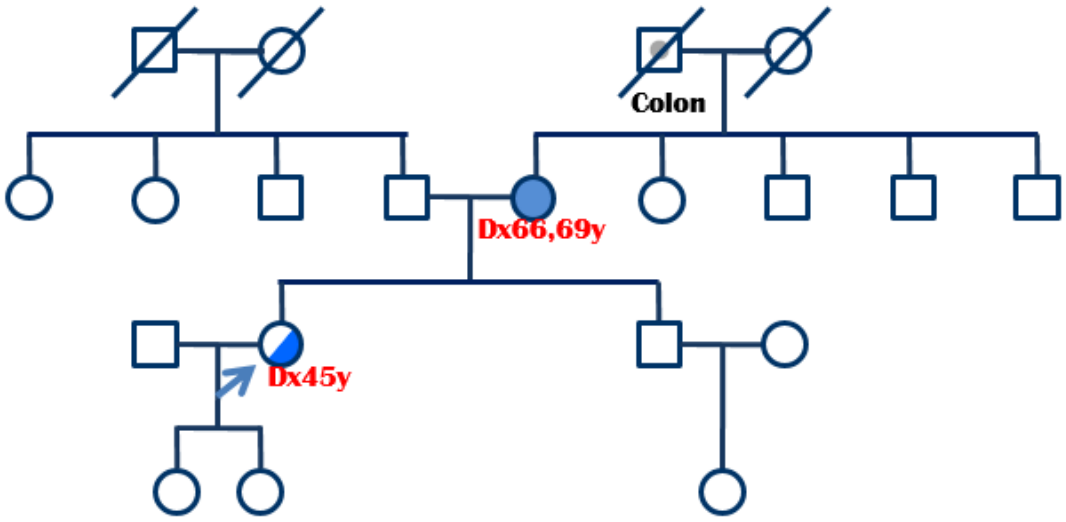
Chromosomal position	Nucleotide change	Protein change	rs number	gMAF EXAC	S	M	P	F	SG	MA	MP	C	PPS	Final	Phenotype and age of diagnosis	Age
73661150	**c.233C>T**	**P78L**	**rs762396676**	**NR**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**91**	**D**	**BC (36)**	
73658791	**c.539G>A**	**R180H**	**rs202162742**	**0,00017**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**T**	**8**	**D**	**BC (45)**	
73626769	**c.1648C>T**	**R550W**	**rs746220717**	**0,00002512**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**78**	**D**	**BC (34)**	
73623552	**c.2926C>T**	**R976W**	**rs752348322**	**0,00003332**	**D**	**T**	**D**	**D**	**D**	**D**	**T**	**T**	**76**	**D**	**BC (39)****	
73659024	c.305C>T	S102L	NR	NR	D	D	D	D	D	D	D	D	88	D		38
73657092	c.929T>A	V310D	NR	NR	D	D	D	D	D	D	D	T	55	D		26
73626311	c.1765C>T	R589W	rs780135537	0,0001422	D	D	D	D	D	D	D	T	66	D		NR
73625270	c.2233C>T	R745W	rs375398949	0,0001	D	T	D	D	T	T	D	T	63	D*		45
73625207	c.2296C>T	R766C	rs186857427	0,000049	D	D	D	D	D	T	D	T	52	D		30

Bold (Cases), White (Controls) NA (Not applicable), NR(Not reported), T (Tolerated), D (Deleterious), * When Predictors were equal (4/4), PPS score was used to discriminate in silico pathogenicity. SIFT (S), MUTTASTER (M), Polyphen-2 (P) information was obtained from Alamut Visual suite version 2.7.2, SNPs&GO (SG) (http://snps-and-go.biocomp.unibo.it/snps-and-go/), FATHMM (F) (http://fathmm.biocompute.org.uk/), MutationAssesor (MA) (http://mutationassessor.org/r3/), MutPred (MP) (http://mutpred.mutdb.org/), Condel (C) (http://bg.upf.edu/fannsdb/). PPS Score was calculated with SNAP2 algorithm (https://www.predictprotein.org/). A higher score assignation (>50) indicated a strong effect induced by point mutations, values (-50<score<50) indicated weak effect and low assignation (<-50) meant neutral/no effect.(**) for this family, the pedigree is not available.

A) Family 2821 CNIO with BC index case, *RECQL5* c.233C>T; P78L



B) Family 2694 CNIO with BC index case, *RECQL5* c.539G>A; R180H



C) Family 2963 CNIO with BC index case, *RECQL5* c.1648C>T;R550W

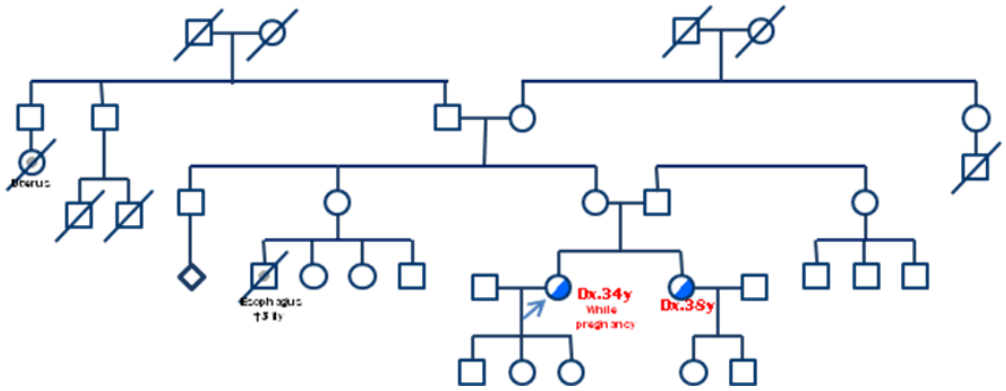
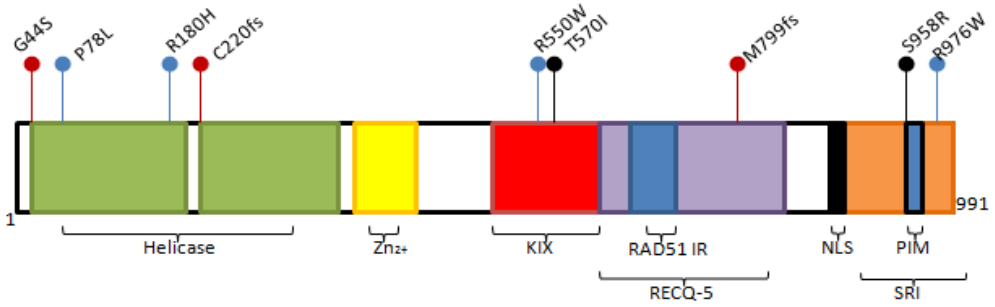
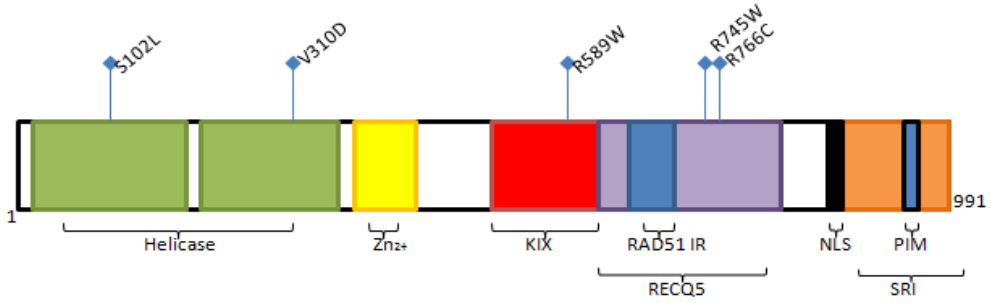


Figure 15. Families harboring *RECQL5* germline missense predicted deleterious variants. Affected members with BC are represented with half colored circles and bilateral BC with complete colored circle; Dx, Age of diagnosis; arrow, index case; other types of cancer are represented with grey dots inside circles. When information was available, the type of cancer is specified in the pedigree.

A) In cases:



B) In controls:



C) Site-directed mutagenesis in functional studies:

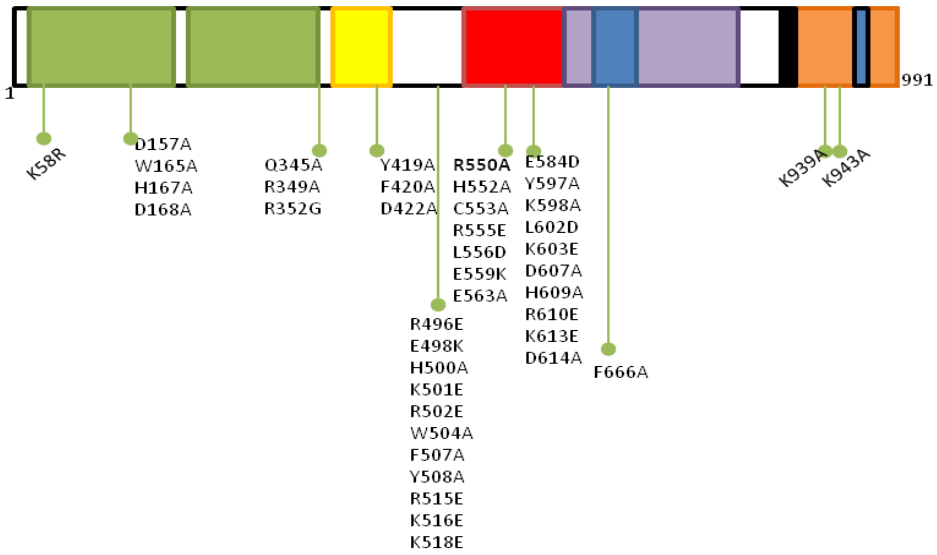


Figure 16. Schematic representation of *RECQL5* with clearly and predicted pathogenic mutations. Domains shown are: Helicase ATP binding site (residues 39-213), Helicase C-terminal (residues 241-354), Zn^{2+} subdomain (residues 365-437), KIX (residues 515-620), RECQ5 (residues 625-829), RAD51 Interacting Region (residues 654-725), NLS (Nuclear Location Signal, residues 866-874), Set2-Rbp1 SRI (residues 900-991), PIM (PCNA interacting region, residues 964-971). Variants originally found by WES are marked with black circles. Clearly deleterious variants are represented with red circles. Frameshift mutations are c.657delC, C220fsX15 and c.2393dupC, M799fsX24. Predicted *in silico* pathogenic variants are shown with blue circles for cases and blue diamonds for controls. Residues with *in vitro* studies are shown in green.

Some variants have been already reported in COSMIC: c.2926C>T, R976W was described as a somatic mutation in leukemia/lymphoma (COSM5706541; Kataoka et al., 2015) and variant c.305C>T, S102L, was found as a somatic variation in stomach cancer (COSM4271256; Kim et al., 2014) and melanoma (COSM4271256; Van Allen et al., 2014). As well, c.2296C>T, p.R766C has been reported in LOVD with no report of function affected, from a gene panel study in healthy controls (no more information regarding gender or age of this individual). C.539G>A, R180H and c.2233C>T, R745W are reported in Exome Variant Server, although no information is available regarding any phenotype associated or data about where the variants were detected.

In silico inference of the missense variants effects in RECQL5 domains was performed in order to establish if there was difference among the variants found in cases and controls.

To compare missense variants in cases/controls located in helicase domain, the predicted effect was compared against the *in vitro* effect of missense variants reported by Newman et al., 2017, who performed site directed mutagenesis of a group of three residues at different positions in the helicase domain: Aromatic loop so called AR-loop, Motif VI in the D2 Subdomain and Zn2+ binding site. Depending on the residue mutated, the missense mutation was able to exert a nule, intermediate or total impairment of the helicase activity. This information helped us to infer the effect of our predicted potentially pathogenic variants.

In a first approach [Figure 17A] the variants located along the subdomain D1 of the helicase (G44S, P78L, R180H) and controls (S102L) were compared against variants in the AR loop (W165A, H167A and D168A). Variant V310D was compared against variants in D2 subdomain (Q345A, R349A and R352G). A401V (negative control) was compared against variants located in Zn2+ domain (Y419A, F420A and D422A). D480G (frequency control), D519A (negative control), R550W (from a case) and R589W (from a control) were evaluated within the KIX domain. Variant T953S (negative control) and R976W (from a case) were evaluated within SRI domain.

Variants acting as negative controls were included in the study for the comparison, these were variants that were ruled out from our original list of missense variants as more than 5 *in silico* predictors assessed a neutral/tolerated effect in the protein, these included variants (A401V, D519A and T953S). The frequency control refers to variant c.1439A>G; p.Asp480Gly (rs820196), with a MAF of 0.25 in the general population. In the present cohort studied for RECQL5, it presented a MAF of 0.1. Interestingly, this is the only missense variant reported in the gene showing a frequency higher than 1% in the general population. Of importance, in

RECQL gene, occurs the same. This would indicate that there is relatively low tolerance for missense mutations in these genes.

In vitro results that were taken into account included increased (ATP)M and Enzyme concentration from Newman et al., 2017. Other parameters that were calculated involved the sites that are modified if in theory the helicase does not function, which are residues located in p.283 (D2 subdomain), E>β-strand (p.160-161), located in subdomain D1 and have a very high impact in the overall function of the helicase. The positions that are predicted to have a severe effect within KIX domain are p.E524 and p.T568 and within SRI domain is p. K928. A score was set in order to measure deleterious (3) or neutral effect (1).

When clustering was performed in order of decreasing deleterious score [Figure 17B], we realized that the RECQL5 variants found in our study in Spanish population grouped together with the highest deleterious score. For the case-variants located within the helicase domain (G44S, P78L and R180H), they resemble the effect of impairment of the helicase shown by variants W165A and H167A and D168A from Newman report and effect in position 283 and 160-161, whereas control-variants located in helicase domains (S102L and V310D) show neutral effect for position p.160-161 which is in contrast with the effect of the case-variants. In addition, predicted effect of V310D is similar to that of negative control (A401V).

Case-variant R550W showed the highest deleterious effect when considering KIX domain uniquely and taking into account comparison against negative control D519A and frequency control D480G. Missense variant R589W found in healthy control had no effect at position E524 and T568, contrasting with variant R550W. As well, missense variants R745W and R766W found in controls located between KIX domain and SRI domain, had a decreased deleteriousness than R550W, with an intermediate effect in KIX position E524 and T568. A high deleterious score was observed for case-variant R976W in SRI domain residue K928, as opposed to negative control T953S.

The last analysis that was performed only took into account the domain for which the variant was located, where we can easily realize the grouping of case-variants (G44S, R79L and R180H) as the most deleterious for helicase domain. As well, as case variants R550W and R976W as the most deleterious in comparison with the variants located in healthy controls that had a neutral effect within the domain where they are located, as in the entire protein [Figure 17C].

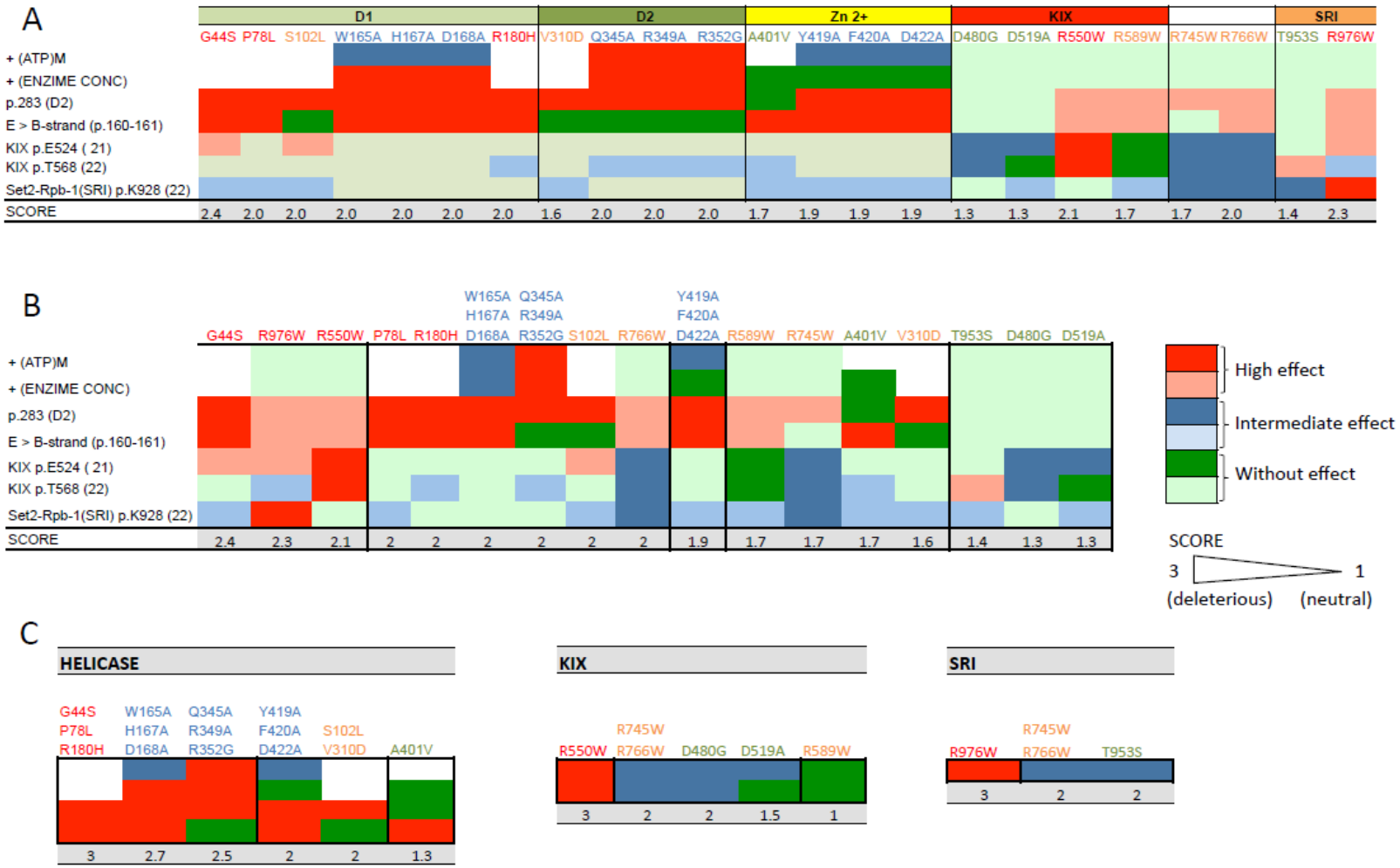


Figure 17. *In silico* inference of potentially pathogenic missense variants in RECQL5 domains.

5.6. Deleterious mutations in the *ATM* gene are present in 1.78 % of breast and/or ovarian cancer Spanish BRCAX families

ATM is already recognized as a moderate BC susceptibility gene, although in Spanish population the reports about the frequency and spectrum of the mutations are scarce. There is solely one report performed in a very small cohort with unselected BC cases, with the caveat of using techniques with low sensitivity (Brunet et al., 2008). Given these, NGS panels were used for *ATM* mutational screening in a cohort of 392 HBOC Spanish BRCAX families looking for pathogenic variants in *ATM* that are related to BC.

Mutational analysis of the *ATM* gene revealed the presence of seven pathogenic mutations in *ATM:* four frameshift, one nonsense, one missense and one splice-site mutation (Table 10). All the variants were found in heterozigosity in index cases. Five mutations were truncating and considered as deleterious. The two remaining, c. 3747-1G>C (Verhagen et al., 2012) and c. 8122G>A (Claes et al., 2013) had been previously classified as pathogenic by functional assays. Other rare variants, some of them potentially deleterious, were found during the screening [Supplementary Table 5], however only mutations unequivocally deleterious were considered for the purpose of this study.

The seven mutations had been previously found in AT and/or BC patients [Table 10]. In Spanish population, c.640delT; p. Ser214Profs*16 and c.8934_8935del; p.Glu2979Alafs*9 have been associated with AT cases (http://chromium.lovd.nl/LOVD2/).

The prevalence of pathogenic germline variants in *ATM* gene was 1.78 % (7 of 392 BRCAX HBOC families in Spanish population). Six mutations were associated with breast cancer only families giving a prevalence of 1.94% (6/308) in this specific group. No mutations in the *ATM* gene were found in a comparative group consisting of 350 index cases from families affected with different diseases not related to breast cancer. These results are consistent with the role of *ATM* as a BC susceptibility gene.

The subsequent analysis of other twenty candidate breast and/or ovarian susceptibility genes revealed deleterious mutations in some of them [Supplementary Table 6]. The most frequently mutated genes following *ATM* were the BC susceptibility genes *FANCM* (Kiiski et al., 2014) and *PALB2* (Antoniou et al., 2014) which were associated with 1.5% and 1.2% of the cases respectively. For the rest of genes the prevalence of mutations was 0.5% at the most.

Table 10. *ATM* germline deleterious mutations and phenotype associated found in 392 Spanish BRCAX families by NGS

Nucleotide Change[1]	Protein change	Gender	Index Case Phenotype and Age of diagnosis	Other cancers In index case	ClinVar	Familial antecedents	References[3]:
c.640delT	p. Ser214Profs*16	F	49 (BC)	No	RCV000169254.1 (2 (Likely pathogenic* Ataxia-telangiectasia syndrome)	Maternal branch: Aunt and cousin also affected with BC before 50y Paternal branch: Two cousins with BC at 52y, 54y	Byrd et al., 1996
c.3747-1G>C	p.?	F	39 (BC)	No	RCV000159717.1 (Pathogenic* Hereditary cancer-predisposing syndrome)	Maternal branch: mother affected with BC and uterine cancer at 60/70y. Aunt with BC at 40y. Grandmother with uterine cancer at 70y	Verhagen et al., 2012
c.3754_3756delinsCA	p.Tyr1252Glnfs*4	M	NR, No diagnosis	Thyroid papilar cancer + Renal Oncocytoma	RCV000164396.1 (Pathogenic Hereditary cancer-predisposing syndrome).	Familial BC and colon cancer	CLINVAR
c.5441delT[2]	p.Leu1814Trpfs*14	F	40 (BC)	No	No	Sister diagnosed with BC at 35y. No other familial cancer antecedents	No
c.6100C>T	p.Arg2034*	F	71 (BC)	Gastric cancer	RCV000122867.2 (Pathogenic* - Ataxia-telangiectasia syndrome), RCV000115222.3 (Pathogenic** - Hereditary cancer-predisposing syndrome).	Maternal branch: Mother affected with BC at 52 y, aunt died of BC at 68y and cousin diagnosed with BC at 60y Paternal branch: two uncles died of colorectal cancer and gastric cancer at 49 and 31y. Two cousins died of pulmonary and prostate cancer, both at 68 y	Telatar et al., 1996
c. 8122G>A	p.Asp2708Asn	F	51/60 BilatBC+OC	No	RCV000132208.2 (Uncertain significance* - Hereditary cancer-predisposing syndrome).	NR	Cavalieri et al., 2006
c.8251_8254del	p.Thr2751Serfs*54	F	<50 (BC)	No	RCV000164777.1 (Pathogenic* - Hereditary cancer-predisposing syndrome).	Cousin and sister affected with BC. Age of diagnosis not available	CLINVAR
c.8934_8935del	p.Glu2979Alafs*9	F	32 (BC)	No	No	Mother and sister with BC before 50y	LOVD

[1]GenBank reference sequence NM_000051.3 with numbering starting at the A of the first ATG, following the HGVS guidelines, www.hgvs.org/mutnomen. [2]Mutation found by WES was not taken in account to establish population percentages. [3] Reference lists the first database or publication informing of the mutation. F (Female), M (Male), BC (Breast Cancer), OC (Ovarian Cancer), NR (Not Reported).

5.7. Search for novel susceptibility genes for hereditary male breast cancer (MBC)

The third part of this thesis is focused in the search of novel susceptibility genes for hereditary male breast cancer.

The MBC family had the following characteristics: presence of three brothers affected with breast cancer (one of them bilateral), absence of other familial antecedents of the disease, no mutations in the *BRCA1/BRCA2* genes. Importantly, in this family the parents were first-degree cousins and there were no women affected with BC in the family. In the literature, there is just another family reported with similar characteristics, which highlights the remarkable phenotype of this family (Wooster et al., 1992) which lead us to explore the X-link and recessive models of inheritance (Figure 7, family 4).

Variant filtering and priorization were performed according to the following criteria selection: X-linked model of inheritance (hemizygous state of the variant in the MBC affected male and no presence of the altered variant in the nephews) or recessive model of inheritance (homozygous state in the MBC affected male, heterozygous for the variant in the nephew whose father was also affected with bilateral MBC and homozygous for the reference or heterozygous in the nephew whose father underwent prophylactic mastectomy). These criteria were taken into account with the information available at the time of the project as well as disposition of the DNA samples.

Candidate variants following a recessive or X-linked inheritance model were identified and validated by Sanger sequencing with specifically designed primers, as well segregation analysis was performed in affected and non-affected members of the family. Initial number of variants after the WES was 16; nonetheless, after segregation analysis, 9 variants were ruled out because they did not fit in our inheritance model. In addition, when the segregation analysis was extended to other male members of the family, variants that did not fit the model were ruled out. The summary of the variants that were finally not included in larger cohorts can be found in Supplementary Table 7. Finally, we selected 7 candidate variants to evaluate in a cohort of 50 hereditary male breast cancers of Spanish population [Table 11], the functions and pathways where the variants are implicated will be discussed further.

Table 11. Candidate variants for hereditary MBC found through WES

Gene	Ch	Chr position	Variant	Protein effect	Functional Class	ID	Model of inheritance	MAF in 50 MBC	MAF ExAc (European)[o]	MAF ExAc (Total)*	Nº Homozygotes ExAc (European)
LIMD1	3	45637230	c.859G>C	V287L	Missense	rs74851793	Recessive	0,01	0,013	0.009	3/121024
TXNDC5	6	7883468	c.1208G>A	R403Q	Missense	rs111331197	Recessive	0,05	0,01	0.007	5/121274
PLCB2	15	40591100	c.749G>C	R250P	Missense	rs199754432	Recessive	0	4,52E-05	6.57E-05	0/120176
GATM	15	45668999	c.88G>A	G30R	Missense	NR	Recessive	0	NR	NR	NR
PTPRH	19	55703094	c.2273C>A	A758D	Missense	rs61734204	Recessive	0,0098	0,007	0.0053	None/120852
YLPM1	14	75265235	c.3235C>T	R1079W	Missense	rs374380781	Recessive	ND	1.5E-04	1.5E-04	None/120616
TKTL1	X	153553774	c.1401+6G>A	p.?	Splice Site Region	rs183407036	X-linked	0	2.29E-04	2.28E-04	4 hemizygotes and no homozygotes from 87657 individuals

NR, No report. ND, not done. Underlined, emphasize genes selected for full coding and exon-boundaries sequencing as short term perspective.*Total include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations. [o]European (non-Finnish)

Of importance, from the results of the Spanish population where we genotyped 50 hereditary MBC cases through Sanger sequencing, we found another hereditary MBC case, who was homozygous for c.1208G>A p.R403Q affecting *TXNDC5* gene. Regarding the clinical history, the patient was diagnosed at 82 years old, diagnosed with an infiltrating ductal BC with familial antecedents (two sisters died due to ovarian cancer at 63 and 75 years old, a sister with Paget BC disease and other relatives with reports of different types of cancer: Colorrectal and lung). There was no information available from the paraffin neither from the tumoral markers.

To assess the implication of these variants in the development of MBC, an analysis was performed with an enlarged cohort of MBC in British population in collaboration with the Institute of Cancer Research (ICR) under the supervision of Dr. Nicholas Orr. Case-control association study included: 1200 cases of MBC and 500 controls (men without the disease and/or familial background of the disease). This study was performed using KASP genotyping technology, using the ABI 7900 instrument. From all KASP assays performed, we did not find other MBC cases where the variants were present in homozygous state for the recessive model of inheritance and in hemizygous state for the X-link model of inheritance. An example of the genotyping results for some of the variants is shown in Figure 18, for variants in *PLCB2* (c.749G>C; p.R250P) and *GATM* (c.88G>A; p.G30R) in 500 British MBC and 500 British healthy controls where we can differentiate clusters formed for homozygous for the reference group (blue), heterozygous (green) and the positive control in a separate cluster (red). Similar results were obtained for the rest of the variants and in 700 extra MBC from British cohort tested, where we detected numerous homozygous for the normal variant, few heterozygotes according to MAF frequency and no homozygous for the recessive model of inheritance.

Considering the rarity of the disease, it is possible that an increased number of cases should be explored in order to increase the possibilities to find an homozygous case for the variant (recessive model) or hemizygous for the variant in X-linked model, but we must remember that the cohort of the ICR is one of the largest collection of MBC available in the world. For increasing the number of cases, it would be necessary to include the variants in a collaborative effort within different consortia.

The presence of the variants was also evaluated in a set of 50 MBC exomes available in the database of the UK laboratory. The variant in *LIMD1* (c.859G>C; p.Val287 Leu) was found in three patients and the variant in *TXNDC5* (c.1208G>A; p.Arg403Gln) was found in one patient. All of them were present in heterozygous state. According to our model, heterozigosity does not confer susceptibility; rather its presence in cases of MBC is due to the frequency of the variants in the population.

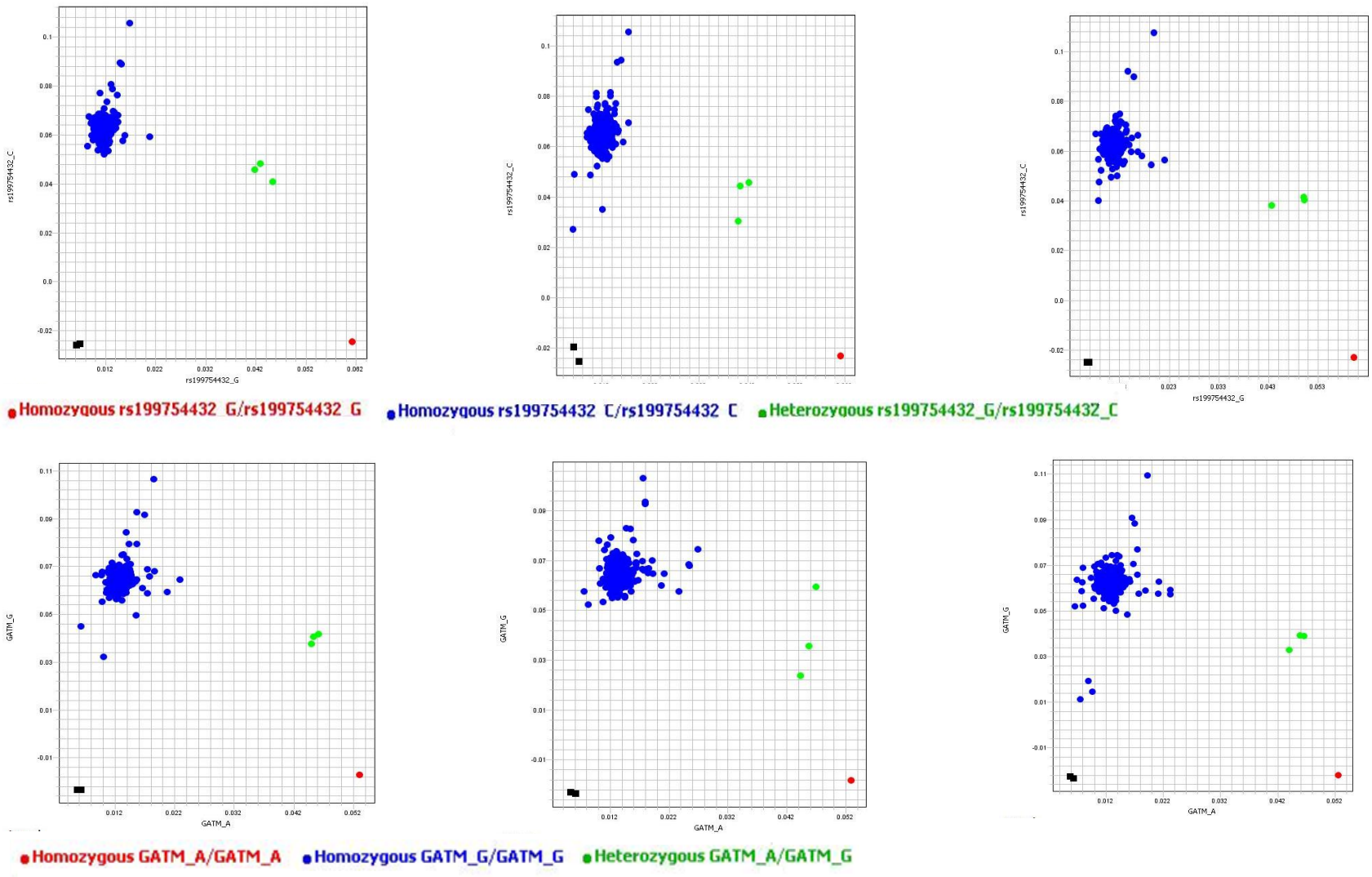


Figure 18. Allelic discrimination plot showing results for variant in ***PLCB2*** (c.749G>C; p.R250P) and ***GATM*** (c.88G>A; p.G30R) in 500 British MBC and 500 British Clusters formed are shown in blue for the homozygous for reference, in green the heterozygous (in both cases, the 3 heterozygous shown in the plots belong to members of the family with known status of the variant) and in red is shown the homozygous for the variant. The black squares correspond to NTC, as internal controls.

6. Discussion

6.1 Exploration of the recessive model of inheritance, genes related with DNA repairing pathways and the dominant model of inheritance in BRCAX families using WES

In this section, we will perform a general overview about the models of inheritance that were explored in the families and the reasons that lead to the exploration from one to another. After, we will discuss the most interesting findings per family.

An example of recessive monogenic inheritance in cancer can be seen in bi-allelic mutations that lead to inactivation of *MUTYH*, which confer higher risk of colorectal cancer (Poulsen et al., 2008). Regarding BC, susceptibility genes had been mostly investigated assuming that they could be inherited as *BRCA1/2*, in a dominant model. Nonetheless, some studies have found by mathematical simulation that other models of inheritance, as the recessive, fit better to explain a percentage of the BC families in which the genetic cause remains not elucidated (Kaufman et al., 2003; Cui et al., 2001; Turnbull et al., 2010; Kuligina et al., 2013). As explained in the introduction, correct selection of families as well as a careful design of the study can potentiate the possibilities of WES to identify novel variants/genes implicated in the disease. That is why we decided to explore the recessive model of inheritance, in this approach, selection of BRCAX families with no previous familial antecedents of the disease and more than two BC siblings affected was prioritized. As an example, variants that affect *TOPAZ1* and *GLE1* as compound heterozygous were found in family 2 [Table 3] which will be discussed further.

No other gene that confers high susceptibility to BC, different than *BRCA1* and *BRCA2* (Foulkes et al., 2008)., has been identified in the last decades, however other genes that confer moderate susceptibility to BC such as *CHEK2*, *PALB2*, *NBN* and *ATM* have been discovered (Foulkes et al., 2013). Remarkably, all of them are implicated in DNA repairing/maintenance pathways. Due to this circumstance, we wanted to explore all the genes with a known function related with DNA maintenance and repairing systems, regardless of the model of inheritance, exploring variants in homozygous, heterozygous and compound heterozygous state (when information was available) and not filtering by impact or *in silico* predictors. Notably, we found a novel pathogenic variant in the well-established BC moderate susceptibility genes *ATM* and quite interesting variants for study in genes that could be potential BC susceptibility genes such as *POLQ* and *RECQL5*, among others [Table 4].

The finding of an *ATM* pathogenic variant linked to BC causality, inherited in a dominant model in family 3, pinpointed that we could not rule out other models of inheritance in the rest of the families. Consequently, we explored this model of inheritance in the rest of the genes.

After case-control association study, variants under the dominant model of inheritance that remained for further analysis were identified in *KANK1*, *CDH15*, *ACTR8*, *CERS5*, *KAT5*, *MTUS1* and *PLEC* for family 1 and *HHAT, KLF15, DEPDC1B, SYNE1, GPRC5B, HBP1, AHNAK* and *RANBP10* for family 2.

In the end, after exploration of different models, our study revealed new genes, which could contribute to breast cancer susceptibility in BRCAX families, all of them family-specific and most probably inherited under a dominant model. These do not imply that the recessive model of inheritance does not exist; rather larger number of families should be carefully analyzed to demonstrate its effect, taking special care in rare families with extreme phenotypes. It has to be noted that the phenotype linked to this type of inheritance is not ample neither easily recognizable, as shown by the initial survey of nearly 2000 families which yielded only 4 candidate families. Contrast in the proportion of families bearing a recessively inherited BC susceptibility allele against a dominant one can be exemplified by Kugilina et al., 2013, whom calculated that 1/250 BRCAX could be potentially associated with the recessive model of inheritance. The numbers contrast when it is compared against a gene that is inherited in a dominant model of inheritance; nonetheless, the recessive model could still explain a percentage of BRCAX families. If we consider the possible incomplete penetrance of the mutations, then this characteristic would make it difficult to recognize the phenotype if there are women who harbor the variant but do not develop the disease.

Another feature that we have to take into account in the families where a dominant variant was found as the most probable causal but variants acting in recessive model were also identified, is the fact that both could add risk to the development of BC under the polygenic model, even though this hypothesis requires further examination.

There are a number of reasons why the use of WES has been questioned as the correct tool for exploring Mendelian or even complex disorders as cancer, being one of them the limited or no coverage of non-coding regions (intronic and intergenic regions), regulatory sequences as promoters, enhancers and microRNAs. Nonetheless, there are an increasing number of diseases for which the causal variant has been found through WES. Also, although WES only explores 1% of the entire human genome, is within this region where major known causal variants fall (Ku et al., 2011). So far, available studies using WES in familial breast cancer have not been conclusive. In the case of the present study, we have confirmed that the use of this technology is justified for the exploration of missing hereditability in BRCAX families by finding novel variants in well-established moderate susceptibility genes associated with the causality of the disease (Family 3),

or novel genes/variants potentially implicated with the development of the disease as *RECQL5* (Family 1 and 2).

6.2 Main findings in BRCAX family 1: Best candidates putatively associated with increased susceptibility to hereditary breast cancer

In this family, under the recessive model of inheritance and after the case-control association study, no variant that fit in this model was classified for further studies. Nonetheless, during the exploration of DNA repairing genes, the main finding in family 1 was a variant located in gene *POLQ:* c.3503T>C, p. I1168T not previously reported [Table 4]. *POLQ* is a gene implicated in repairing DNA double strand breaks (DSBs) by the error-prone micro homology mediated end joining pathway (Kent et al., 2015). It also participates in translation synthesis and DNA replication (Goullet de Rugy et al., 2016). Interestingly, it has been demonstrated that codepletion of *BRCA2* and *POLQ* increase sensitivity to cisplatin-resistant lung cancer cells (Dai et al., 2016), the characterization of both the mechanism and the effect of this synthetic lethality phenomenon in BC would be of interest for personalized medicine.

Under the dominant model of inheritance, variants that remained for further studies are located in *KANK1*, *CDH15*, *ACTR8*, *CERS5*, *KAT5*, *MTUS1* and *PLEC*. From these, two seem to be quite interesting as they are frameshift mutations that lead to a truncated product in two different genes: c.3709del, p.A1237fsX11 in *KANK1* and c.1316del, p.P439fs-X38 in *CDH15* [Table 6].

KANK1 (9p24.3) encodes a product of 1352 residues and has a major role in cytoskeleton formation by regulating actin polymerization, inhibits actin fiber formation and cell migration and inhibits RhoA activity, participates in establishment and persistence of cell polarity during directed cell movement in wound healing (Sarkar et al., 2002; Kakinuma et al., 2009). It has been proposed as tumoral supressor in renal cell carcinoma, as its depletion or loss leads to cell movement and invasion. Expression of the gene was found reduced either by loss of heterozigosity and/or methylation at CpG sites in this gene in renal carcinoma samples (Sarkar et al., 2002; Roy et al., 2005). In addition, its depletion induces hyperactivation of RhoA, centrosomal amplification and genomic instability, a hallmark of cancer (Suzuki et al., 2017).

CDH15 (16q.24.3) encodes a member of the cadherin-protein family, which are calcium-dependent intercellular adhesion glycoproteins. Other members of this family have been implicated in BC, as *CDH1* which is a gene that confer high susceptibility to this disease within cancer related syndromes (Pharoah et al., 2001). *CDH15* is a potential tumor suppressor as LOH analysis performed on 62 mammary tumors induced in (BALB/c x C57BL/6) F1 mouse

mammary tumor virus/neu transgenic mice showed preferential LOH of the region containing this gene (Cool et al., 1999), an observation that was seen also in human sporadic breast cancer tumors (Kremmidiotis et al., 1998).

KAT5 (c.291C>G, D97E), also known as TIP60, has remarkable functions, all related with DNA repair and *BRCA1*, which turns this gene into an excellent candidate. It encodes an acetyltransferase involved in apoptosis, cell cycle and DNA damage response. Tip60 is involved in homologous recombination, DSB repair and its deficiency reduces BRCA1 at DSB; consequently, impairing HR and conferring sensitivity to PARPi (Tang et al., 2013). It is considered as a novel breast cancer tumor suppressor gene required for breast cancer progression (Bassi et al., 2016). It can inhibit metastasis by regulating epithelial-mesenchymal transition (Zhang et al., 2016).

A missense variant was located in *MTUS1* gene (c.3554A>C, p.N1185T). This gene (Microtubule associated tumor suppressor) has been described in different types of cancer, where loss of function or reduced mRNA levels (e.g. by negative coregulation among microRNAs) are related with worsened prognosis and higher invasive capability of the tumors (Gu et al., 2017; Kara et al., 2016; Ozcan et al., 2016; Zhao et al., 2015). In familial BC, it inhibits BC cell proliferation, delays the progression of mitosis by prolonging metaphase, and reduces tumor growth (Rodrigues-Ferreira et al., 2009; Frank et al., 2007).

It was interesting to find that a variant (c.12886C>T, p. R4296C) affecting *PLEC* gene, which encodes a protein of 4684 aminoacids, that has different roles, as structural protein it maintains cell, tissue integrity, remodelates cytoarchitecture, cell shape and also acts as scaffold protein for assembly, positioning, and regulation of signaling complexes (Wiche et al., 1982; Liu et al., 1996). More interestingly, PLEC interacts with *BRCA2* through an interaction that controls the position of the centrosome, if this interaction is dissociated it leads to centrosome disorganization, genomic instability and cancer development (Niwa et al., 2009). Regarding the phenotype of cases found in Openarray platform bearing this variant: 8/12, 66% were breast cancer-only cases. The presence of this variant in a high number of cases, as well as its interaction with BRCA2, turns this gene into a quite interesting gene for full coding sequencing in Spanish population, to explore the association with increased susceptibility to BC.

6.3 Main findings in BRCAX family 2

6.3.1. *RECQL5*, another DNA helicase potentially involved in increased breast cancer susceptibility

While studying the genes related with DNA repairing systems, we found two variants affecting *RECQL5* (c.1709C>T, p. T570I and c.2874C>G, p.S958R). There are important features why RECQL5 is an excellent candidate for deep study, as: members of the RECQL helicase family of proteins, as BLM, WRN, and RECQL4, are related with diseases that predispose to cancer development, e.g. Bloom, Werner and Rothmund-Thomson syndromes (Ellis et al., 1995; Goto et al., 1996; Lindor et al., 2000; Bohr et al., 2008). From all the members of the family, RECQL5 is the only one that has not been reported in any genetic condition. As mentioned previously, RECQL5 participates in a plethora of cellular activities: DNA repairing, recombination, replication, transcription, genome stability, homologous recombination and cell survival (Croteau et al., 2014). It has been shown that *RECQL5* suppresses tumor formation and *RECQL5* knock-out mice have increased cancer predisposition (Hu et al., 2007; Hu et al., 2010). In addition, *RECQL*, another member of REC-helicases family, has been recently identified as a new putative BC susceptibility gene (Cybulski et al., 2015; Sun et al., 2015).

Given all the characteristics that *RECQL5* shares with known high/moderate BC susceptibility alleles, we decided to study all the coding regions and exon-intron boundaries of the gene using Illumina TruSeq Custom Amplicon platform in a large series BRCAX BC-only cases and controls (1463). From the mutational screening, we found three germline unequivocally pathogenic variants in cases, one missense (c.130G>A, G44S) and two frameshift (c.657delC, C220fsX15; c.2393dupC, M799fsX24) that led to a premature truncated protein, this would represent 0.42% of the cases. In controls, no clearly deleterious mutation was found.

c.130G>A, G44S is a novel missense mutation that is predicted to affect nearest splicing site leading to the loss of several exons still to be determined, which leads to truncation of almost 95% of the complete product. Also, this variant introduces an extra helix in the domain and loss of catalytic residue N45, according to MutPred. c.657delC, C220fsX15 leads to the loss of important domains for helicase activity (ATP-binding site and Helicase C' terminal) as well as all domains located further (Zn^{2+} subdomain, KIX, RECQ-5, RAD51 IR and SRI domains). c.2393dupC, M799fsX24 truncates the C'terminal extension of RECQL5, which includes PIM and SRI domains.

Segregation analysis performed to sisters affected with BC from index case harboringc.130G>A, G44S revealed incomplete segregation of the variants which suggest that the

variant could be associated with moderate risk. Unfortunately, it was not possible to extend segregation analysis in other members of the families from BC index patients harboring c.2393dupC, M799fsX24 and c.657delC, C220fsX15.

Mean age of onset at first diagnosis of BC in patients with *RECQL5* deleterious mutations in Spanish population was 36.7 years, which is earlier than BC diagnosis for *BRCA1*/*BRCA2* mutation carriers (being 43.6 and 42.8, respectively) and their sporadic counterparts in Spanish population (Diez et al., 2003). Average age of BC first diagnosis in patients carrying *RECQL* deleterious mutations was 45.1 years in Chinese population (Sun et al., 2015), 48.9 years for French-Canadian and 54.5 years for Polish population (Cybulski et al., 2015), the other RECQ-helicase which has been related to increased susceptibility to BC. Nonetheless, there are two facts to consider regarding this data: very few cases of BC patients with *RECQL5* deleterious mutations were taken into account for this value and the mean age of first BC diagnosis was 39.3 years in the Spanish BRCAX cases cohort studied (n=700).

In the majority of the cases where we detected *RECQL5* germline pathogenic variants, complete information about the IHQ of the tumor was not accessible. For c.2874C>G, S958R it was shown in paraffin embedded tumor, that IHQ was ER+, PR+. For c.2393dupC, M799fsX24, both bilateral cases were triple negative. No more information from markers was available from c.130G>A, G44S and c.657delC, C220fsX15. This data would be of interest for classification, as a considerable percentage of BC cases with BC *RECQL* germline pathogenic variants were ER+, PR+, Her2 negative, resembling *BRCA2* associated cases (Sun et al., 2015).

Bearing in mind that all the *RECQL5* germline pathogenic variants described in this study where found in heterozigosity, haploinsufficiency and/or inactivation of wild-type allele by a second hit could be associated with BC development, although the molecular mechanism has to be established to validate this hypothesis. It has been observed that knockout mice for RECQL5 do not show loss of heterozigosity, which suggest that LOH alteration is not the primary underlying mechanism of cancer susceptibility (Hu et al., 2007), rather it could be related with the hyper recombination favored by the absence or partial function of RECQL5 if mutated which would increase genome instability and gross chromosomal changes that tend to be oncogenic. Other mechanisms could be associated taking into account the participation of the enzyme in the relief of replication stress generated by stalled DNA forks (Di Marco et al., 2017) and the halting of transcription when DNA double strand breaks are detected (Islam et al., 2010). In addition, the antagonic role of RECQL5 in disruption of the RAD51 presynaptic filaments (Schwendener et al., 2010) against BRCA2 that promotes assembly of this latter could be deregulated in an oncogenic condition [Figure 19].

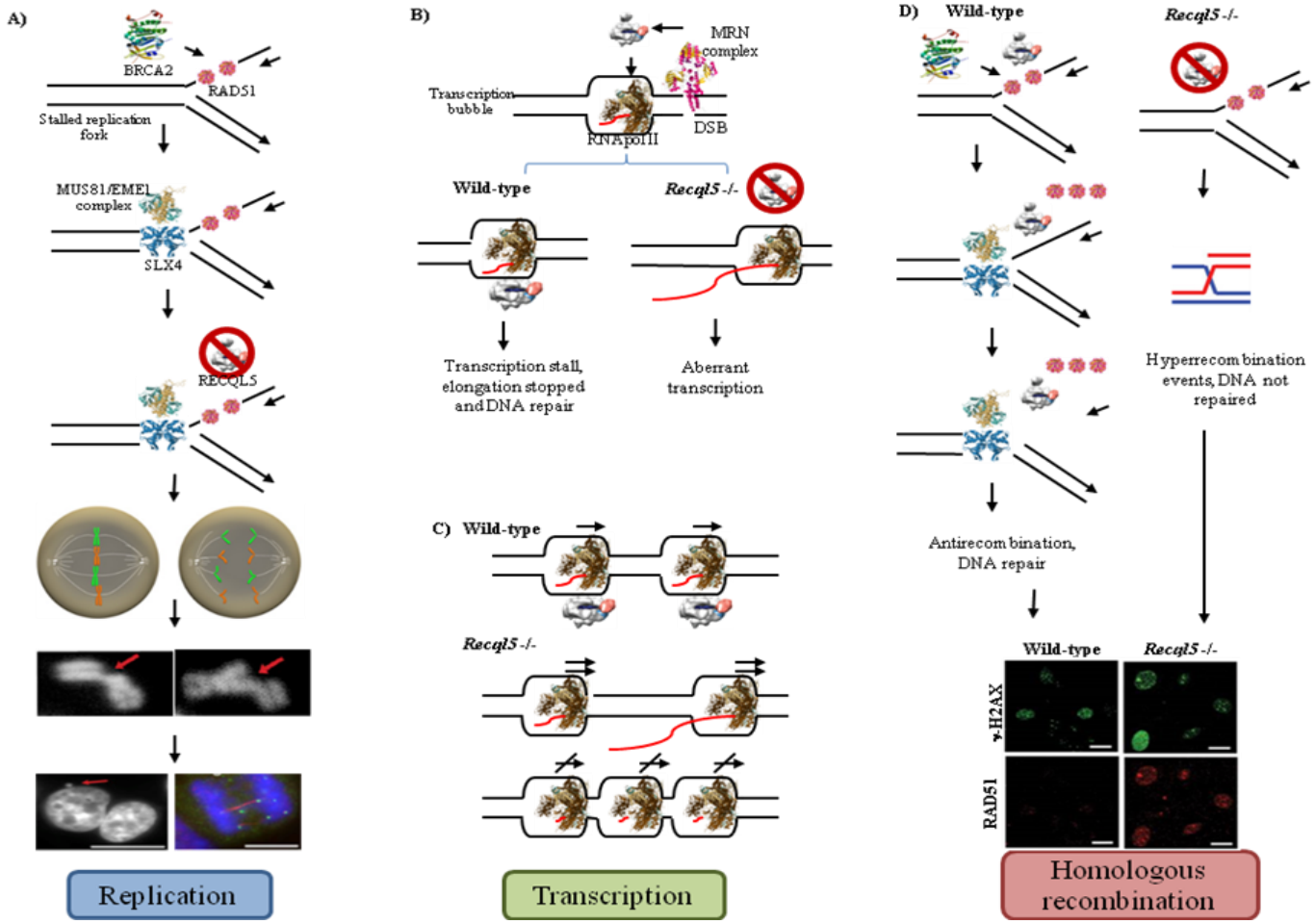


Figure 19. General overview of potential mechanisms driving to tumorigenesis caused by absence/partial expression/functional impairment of RECQL5. A) In replication, absence/partial function of RECQL5 leads to inhibition of fork cleavage due to non disruption of RAD51 filament and no recruitment of MUS81 complex which, after mitosis, results in genomic instability and cells that present aberrant micronuclei and aberrant chromosomal segregation. B) Dual mechanism of RECQL5-transcriptional repression: MRN complex recruits RECQL5 inhibiting transcription by steric block to prevent PolII toward the site of the DBS and inhibition of elongation at stall site. C) In transcription, at elongation phase RECQL5 is not only relieves transcription stress, but also regulates the movement of RNApolII across genes. Without this enzyme, there is an increase of stalling, pausing and/or backtracking of RNApolII. D) In Homologous Recombination, antirecombinase activity of RECQL5 leads to disruption of RAD51 filament to prevent aberrant HR. Impaired or nule RECQL5 activity leads to hyper-recombination and higher numbers of RAD51 foci (Adapted from Di Marco et al., 2017; Hu et al., 2007; Kassube et al., 2013; Schwendener et al., 2010; Saponaro et al., 2014).

By using a combination among *in silico* predictors and predict protein scores, we classified potentially deleterious missense germline variants in cases (c.233C>T, P78L; c.539G>A, R180H; c.1648C>T, R550W; 2926C>T, R976W) and controls (c.305C>T, S102L; c.929T>A, V310D; c.1765C>T, R589W; c.2233C>T, R745W; c.2296C>T, R766C). From these, it is important to mention that the majority of the predicted germline pathogenic variants are highly conserved residues among species centered in the RECQ-helicase, KIX or SRI domains [Figure 17].

One of the variants with very high *in silico* predicted pathogenicity was c.1648C>T, R550W, present in a case with BC diagnosed at 34 years old. Site directed mutagenesis of residue R550 has been shown to impair protein folding and abolish interaction of POLR2A (Islam et al., 2010). In an analogous way, variant c.1648C>T, p.R550W is predicted by MutPred to introduce structural changes (gain of a helix and loss of a loop) that may lead to incorrect folding of the protein. Also due to its location, at nucleotide 4 at the beginning of the exon 13, the variant is predicted to affect nearest splicing site.

An *in silico* study was performed centered in the domains of the protein to study the effect of missense variants and to compare between cases and controls. Positive controls were taken into account considering functional studies published by Newman et al., 2017 were point mutation lead to null (Y419A, D168A), partial (H167A, D422A, F420A) or total decrease (R349A, Q345A, W165A and R352G) in the helicase activity of RECQL5 depending in the position of the variant inside of the helicase domain. The negative control used for computational comparison included missense variants with a tolerated/neutral effect found in the Spanish population and variant c.1439A>G; p.Asp480Gly (rs820196) quite frequent in our cohort and a GMAF in EXAC of 0.2 in general population. Of importance, no other missense variant have been reported in EXAC with GMAF higher than 1% which points to the fact that missense variants are not well tolerated in this gene. The results from these *in silico* approach showed that the missense pathogenic variants detected in cases exert a putative impairment in the helicase functionality, which is not seen in variants detected in controls for variants detected within this domain [Figure 17]. As a perspective, functional studies are planned in order to differentiate the effect of missense variants at helicase level.

For variants located in KIX domain, it was found that the most deleterious was present in case R550W, for which there are *in vitro* studies proving its pathogenicity (Islam et al., 2010), whereas variants located nearby this domain and present in controls have an effect similar to that one from frequency and neutral missense variants. In a comparable way, in SRI domain variant R976W from case has an increased deleterious effect in comparison with controls. As KIX and SRI domain interact with RNA-Polymerase II, by halting transcription, and then the capability of

binding to this enzyme could be tested *in vitro*. It has been proposed that KIX and SRI domain have different roles at two stages of the transcription. SRI domain is more implicated in the elongation stage, whereas KIX domain in recognition of initiation complex of transcription (Aygün et al., 2009; Aygün et al., 2010). As well, they bear distinct capabilities to recognize hyper/hypo phosphorylated form of RNA-poll II, for which specific assays could be designed in order to test whether or not the different mutations affect at distinct levels of the transcription process, as well as the different phosphorylated forms of the enzyme. As a perspective, proliferation assays, apoptosis assays and interaction with RAD51 and sensitivity to a panel of drugs (including camptothecin derivatives) to study mechanisms of synthetic lethality which could be affected due to the presence of missense pathogenic variants when compared with the WT RECQL5 could also be performed in order to get insights about the connection between the predicted pathogenic variants and the development of BC.

From the *in silico* inference of missense variants in RECQL5 domains, computational approach suggest that missense variants located in cases are, in fact, pathogenic germline variants, which would translate to an increase of the prevalence (7/700). This data, if *in vitro* corroborated, would mean that RECQL5 germline pathogenic variants would confer a moderate/high risk in 1% of all BRCAX BC-only cases, a proportion similar to the explained by other moderate-high susceptibility BO/OC genes.

Finally, alike tumors harboring *BRCA1* or *BRCA2* germline mutation with increased and selective sensitivity to PARP inhibitors (Farmer et al., 2005; Lord et al., 2017), which is now the cornerstone of personalized therapy for these type of patients, it has been shown that *RECQL5* deficient cells are hypersensitive for treatment with Topoisomerase I (camptothecins, CPT) in colorectal cancer cell lines (Wang et al., 2011). Nonetheless, they are resistant to cisplatinum, mitomycin C and etoposide (Hu et al., 2009). Thus, for carriers of *RECQL5* germline pathogenic variants already affected with BC, this knowledge could guide the selection of specific chemotherapeutic regimens for RECQL5 defective BC cancers (e.g. derivatives of CPTs, irinotecan and topotecan). In normal conditions, accumulation of CPT-DNA-Top1 lesions, would lead to activation of DNA damage response, which includes RecQ5-helicase (Shamanna et al., 2016). Nonetheless, if the RecQ5-helicase is missing, then selective elimination of this type of tumor would happen through a hitherto not described mechanism of synthetic lethality.

Missing heritability is becoming unraveled with NGS technologies. This study supports the use of WES to identify new genes involved in BC susceptibility. The prevalence of at least 0.42% *RECQL5* pathogenic germline variants associated with BC in Spanish population, justifies deeper attention to this gene. To calculate the exact risk, larger cohorts need to be evaluated and *in vitro*

studies are desirable to verify pathogenicity of undetermined missense variants predicted to be highly deleterious by numerous computational methods. Further exploration is justified to guide prevention, early diagnosis and future treatment alternatives in carriers of *RECQL5* germline pathogenic variants.

6.3.2. Other findings in family 2 exploring dominant and recessive model of inheritance

In family 2, under the dominant model of inheritance, we found 13 candidate variants. After case-controls association studies, variants that remained for further insights are located in genes: *HHAT, KLF15, DEPDC1B, GPRC5B, SYNE1, HBP1, AHNAK* and *RANBP10*.

From these, one of the most interesting variants is located in *HHAT* c. 1096_1097del, p.L366fsX86 that generates a truncated product, that losses important functional domain beyond residue 366. *HHAT* (1q32.2) encodes for Hedgehog acyltransferase, catalyzes the transfer of the fatty acid palmitate into Sonic Hedgehog (a pathway implicated in BC) (Konitsiotis et al., 2015). Strongest evidence for selection in further studies was a report where depletion of HHAT decreased selectively anchorage-dependent and anchorage-independent proliferation of ER positive, HER2 amplified and tamoxifen-resistant cells but not triple negative breast cancer cells lines, which is of interest in the treatment of BC patients harboring mutations in this gene (Matevossian et al., 2015).

Another two interesting missense variants are located in *KLF15* (c.965A>G, p. K322R) and in *DEPDC1B* (c.739A>G, p. M247V). The first one encodes (3q21.3) a protein named as Kruppel like factor 15, is a transcriptional regulator involved in various biological processes, including cellular proliferation, differentiation and death. In BC, it is a putative tumor suppressor, which partially stalls cancer growth through p21 up-regulation. Presence of this protein was associated also with clinicopathological factors predicting a better clinical outcome (Yoda et al., 2015). Downregulation of KLF15 by miR-262 leads to proliferation and invasion in human breast cancer cells (Wang et al., 2017). Interestingly, in uterus KLF15 regulates negatively the β-estradiol-induced epithelial cell proliferation by inhibition of DNA replication licensing which is indicative of a role at two levels: inhibition of hormonal effect and DNA maintenance system, which make this gene a candidate susceptibility gene in HBOC (Ray et al., 2012). In the case of the other gene, *DEPDC1B* (5q12.1), it encodes a protein of 529 residues that is involved in cell migration, intracellular signal transduction, positive regulation of Wnt pathway and regulation of small GTPase signal transduction (Marshesi et al., 2014). In BC, it was found that tumors from *BRCA1* mutation carriers show LOH of chromosome 5 (Johannsdottir et al., 2006). Homozygous 5q deletions spanning 700 kbp which included *DEPDC1B*, had been reported twice by different

studies in tumors from a *BRCA1* mutation carriers (Johannsdottir et al., 2006; Jönsson et al., 2005).

HBP1 (8q24.3) encodes a transcriptional repressor with a HMG box DNA-binding domain that binds to promoter region of target genes. Plays a role in the regulation of the cell cycle, senescence and in inhibition of the Wnt pathway (Sampson et al., 2001; Zhang et al., 2006). The molecular mechanism in Wnt signaling arises from the inhibition of transcriptional activation of Wnt signaling (*cyclinD1* and *c-MYC*). For this latter, HBP1 interacts directly to c-MYC and negatively regulates its transcriptional activity (Escamilla-Powers et al., 2010). It can interact with RB1 and enhance binding to the $H1^0$ promoter. Disrupts DNA interaction with TCF4 (Lemercier et al., 2000). A report in BC found *HBP1* mutants/variants are present somatically in tumoral samples of invasive ductal carcinoma (IDC); interestingly, in this type of tumor (IDC) even without somatic mutations of *HBP1*, the levels of *HBP1* mRNA were reduced, which was associated with Wnt signaling increased tumorigenic proliferation and invasiveness. This resembles the mechanism exerted by miR-17-5p that suppresses HBP1 and consequently enhances proliferation and invasiveness (Li et al., 2011). Noteworthy, low/null levels of *HBP1* mRNA correlates with bad prognosis as well as BC recurrence (Paulson et al., 2007). Concerning phenotype, it was interesting to find variants with numerous cases associated with breast cancer only patients, as for *HBP1* c.1167A>C, p.Q389H in 11 cases, from which 81% were BC-only.

AHNAK (11q12.3) encodes a protein of 700KDa, which is a structural scaffold protein that participates in cell structure and migration, cardiac calcium channel regulation, and tumor metastasis (Shtivelman et al., 1992). An interesting role in DNA maintenance complexes as Ahnak can interact with DNA ligase IV-XRCC4 complex in DNA non-homologous end-joining and stimulates DNA ligase IV-mediated double-stranded ligation (Stiff et al., 2004). In BC, it has been reported overexpression of Ahnak resulted in growth retardation and cell cycle arrest through downregulation of c-Myc and cyclin D1/D2, whereas depletion of Ahnak leads to significantly progressed hyperplasia of mammary glands in a mouse model and in human BC tissues its expression is lower in tumoral vs normal tissue(>50%). This information suggest that *Ahnak* is a novel tumor suppressor that functions via modulation of TGFβ/Smad pathway (Lee et al., 2014). Ahnak has been localized in extracellular micro vesicles that allow mammary carcinoma cells to communicate with surrounding cells from its niche and transform them (Silva et al., 2016). Concerning phenotype, variant located in A*HNAK* (c.7964G>T, p.G2655V) was found in 6 cases, with 4 of them being BC only cases.

Regarding the recessive model of inheritance, a quite interesting gene where potential pathogenic variants were found as compound heterozygous was *FBF1*. Two variants were

inherited from the mother (c.770C>G, p.P257R and c.1808G>A, p.R603Q) and one from the father (c.2039C>T, p.A680V). The three are present in the affected sister with BC, in exon 11, 18 and 19, meanwhile they are WT in the healthy sister. This gene encodes a protein known as FAS binding factor 1 or Albatross. It has not been previously associated to BC cases, nonetheless it is implicated in epithelial cell polarization (in the apical junction complex (AJC) (Sujimoto et al., 2008) and regulation of PLK1 activity at G2/M transition. Also, it is a protein that participates in the union of Fas antigen (TNFRSF6), a cell death mediator. FBF-1 shows sequence similarity with plectin which is a structural protein with a potential role in BC that we found while studying other models of inheritance and will discuss further (Schmidt et al., 2000).

Although we consider that the variants affecting *RECQL5* could be associated with the causality of the disease in this family, still other genes found under the dominant and recessive model of inheritance are of interest for deeper molecular and statistical exploration, as they may be acting as modifiers of the risk and could influence at some extent the BC development in a family-specific way, although this hypothesis require further exploration.

6.4 Main findings in BRCAX family 3: Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the *ATM* gene

Family 3 was an example in which we were able to find the causal variant in an already known moderate BC susceptibility gene using WES. We found a variant located in *ATM* (c.5441delT; p.L1814WfsX14) that leads to a truncated product and provided evidence of the causality of the disease. Due to this finding we decided to perform complete coding sequencing of the gene in a large series of cases and controls to explore frequency and spectrum of *ATM* mutations associated with HBOC in Spanish cohorts due to the lack of this type of information in this population (Brunet et al., 2008; Graña et al., 2011).

By using NGS panels, we determined a prevalence of *ATM* pathogenic germline variants of 1.78 % (7 of 392 BRCAX HBOC families), which is slightly higher than that reported in other populations which varies between 0.69% in Japanese population (Hirotsu et al., 2015) to 1.41% in US population (Kurian et al., 2014) in BRCAX families. No mutations in the *ATM* gene were found in a comparative group consisting of 350 index cases from families affected with different diseases not related to breast cancer. Although this is not strictly a control group because patients were not free of disease, we think that the comparison is valid for the purpose of this study as we did not find any deleterious mutation, and we assume that the prevalence in the general population would be even lower.

Regarding the phenotype, the majority of mutations were found in index cases affected with BC (5/7) cases. Only the mutation (c.8122G>A, p.Asp2708Asn) was present in an index case with bilateral BC and ovarian cancer. Increased risks for bilateral BC occurrence among *ATM* heterozygous carriers have been observed (Broeks et al., 2000). Index case harboring c.6100C>T; p.Arg2034* mutation presented breast and gastric cancer and index case carrying c.3754_3756delinsCA, p.Tyr1252Glnfs*4 developed thyroid papillary cancer and renal oncocytoma [Table 10]. This was not surprising, as adult *ATM* heterozygous carriers can develop solid tumors which include cancers of the digestive track, lung and gallbladder, although with lower frequencies in comparison with BC (Taylor et al., 2015).

Segregation analysis of family members from index case harboring c.3754_3756delinsCA, p.Tyr1252Glnfs*4, revealed two mutation carriers diagnosed with BC at 43 and 45 years, which strengthens the role of *ATM* pathogenic mutations in the development of BC. Importantly, average age of onset at first diagnosis of BC in patients with *ATM* deleterious mutations in Spanish population was 47.42 years, which is amid BC diagnosis for *BRCA1*/*BRCA2* mutation carriers (being 43.6 and 42.8, respectively) and their sporadic counterparts in Spanish population (Diez et al., 2003).

The prevalence found, almost 2% of *ATM* pathogenic variants in Spanish BC only families which is higher than the reported in other populations, justifies considering its incorporation in routine genetic tests in clinical setting for BRCAX families in Spanish population to detect *ATM* heterozygous carriers whom could benefit from specific monitoring programs.

6.5 Main findings in BRCAX family 4: Novel susceptibility alleles potentially implicated in hereditary male breast cancer

MBC accounts for a total of 1% of all the breast cancer cases (Speirs et al., 2009). From all MBC, 10% are hereditary and among 60 – 70% of these cases are caused by pathogenic mutations in *BRCA2* gene (Tournier et al., 2004). Nonetheless, little is known about other MBC susceptibility genes. In addition, clinical and pathological characteristics of MBC do not overlap female BC, mortality and survival rates have not improved significantly and its incidence is increasing (Rizzolo et al., 2013). Little is known about other susceptibility genes implicated in the development of the disease that could guide prevention and measurements for advanced cases (Giordano et al., 2004).

By using WES, in an exceptional MBC BRCAX family with an apparent recessive or X-linked model of inheritance [Figure 7, family 4], we selected candidate variants that could be related

with the disease [Table 11]. While exploring the presence/absence of the variants in a cohort of MBC Spanish cases with familial and personal background of the disease, we have found another MBC case who was homozygous for c.1208G>A p.R403Q affecting *TXNDC5* gene which has been related to Androgen Receptor. A large percentage of MBC (approximately >60%) are related with somatic mutations occurring in hormone signaling pathways (Sas-Korczynska et al., 2015),. Of transcendence, germline mutations in the androgen receptor have been found associated with other cases of multiples MBC within the same family (Wooster et al., 1992), as well as related with hormone driven tumors due to dysregulation in the hormone metabolism of androgens (Wang et al., 2015), as prostate cancer (Baumann et al., 2012). *TXNDC5* encodes a member of disulfide isomerase family of proteins, acting as a chaperone of endoplasmic reticulum. As a result, *TXNDC5* interacts with many cell proteins, contributing to their proper folding and correct formation of disulfide bonds through its thioredoxin domains. In addition, it is a cellular adapter, playing an important role in cellular physiology. Deregulation of this protein has been associated with different pathologies like arthritis, cancer, diabetes, among others (Horna-Terrón et al., 2014). A recent paper has described an interaction between *TXNDC5* and AR, to increase its stability and thus, enhance its transcriptional activity (Wang et al., Oncogene 2015). Noteworthy, *TXNDC5* SNPs could be related to the development of hepatocellular carcinoma tested in a cohort of Korean males (Park et al., 2013). Although the interaction between TXNDC5 and AR is not completely understood, in our case, having found two MBC cases that are homozygous for the same variant is quite interesting taking into account the rarity of the disease and the pathway implicated. Further insights are desired as *TXNDC5* could be a MBC susceptibility gene and for the design of personalized medical regimens in MBC (Di Lauro et al., 2015).

We have selected variants with a recessive model of inheritance affecting other quite interesting genes like: *LIMD1*, *PLCB2* and *YLPM1*, among others, prioritized due to the functions and pathways where they are implicated.

Missense variant affecting *LIMD1* is c.859G>C, p.V287L, a gene (3p21.31) that plays multiple roles: assembly of numerous protein complexes and is involved in several cellular processes such as cell fate determination, cytoskeletal organization, repression of gene transcription, cell-cell adhesion, cell differentiation and migration (Bai et al., 2011). It is also a negative regulator of Hippo signaling, a pathway that is implicated in mammary gland development and male BC (Das Thakur et al., 2010; Pinto et al., 2014). Interestingly, LIMD1 interacts with BRCA2 in a well conserved region (amino acids 2,750-3,094), by disrupting this interaction, centrosome localization of BRCA2 is affected yielding abnormal cell division (Hou et al., 2016). This latter is of remarked importance, as BRCA2 is the only gene of susceptibility to MBC which is associated with a large percentage of hereditary cases.

Missense variant affecting *PLCB2* (c.749G>A, p.R250P), a gene (15q15.1) encodes a Phospholipase C Beta 2, an enzyme needed for production of the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3). PLCB2 is highly expressed in advanced female BC and is associated with poor outcome (Bertagnolo et al., 2006). May improve some malignant characteristics of tumor cells, like motility and invasion capability and may influence the modification of cell shape that characterizes cell division, motility and invasion (Bertagnolo et al., 2007).

Missense variant affecting *YLPM1* is c.3235C>T, p. R1079W, a gene (14q24.3) related to telomere maintenance, by reducing telomerase activity during differentiation of embryonic stem cells by binding to the core promoter of TERT and controlling its down-regulation (Armstrong et al., 2004). It is a candidate gene for mammary gland development in pigs, although further characterization will be needed in humans (Verardo et al., 2016).

Regarding the X-linked inheritance model, we have one candidate variant affecting *TKTL1* gene (Xq28), which is a splice site variant. This gene is related with the pentose-phosphate pathway, with the glycolytic pathway and related with tumor metabolic reprogramming (Diaz-Moralli et al., 2016). It promotes cell proliferation and metastasis, is a marker of certain tumor tissues and induced by hypoxia (Földi et al., 2007). In BC, its expression was higher in BC and DCIS in comparison with normal tissues (Schmidt et al., 2010).

Even though our findings in the case-control association were negative for the British cohort (we did not find any other variant homozygous for the recessive model and hemizygous in the X-link model), still we have to consider that these findings do not imply that the variants can be ruled as potential MBC susceptibility genes; instead, larger number of MBC cases should be explored.

In summary, these set of studies using Whole Exome Sequencing for exploring BC missing hereditability has guided to the exploration of different models of inheritance, and lead to the discovery of causal variants family specific and other variants, which overall, could be acting jointly to increase the risk of BC. Although there is still work to do in light of statistical power and functional characterization of potentially deleterious variants, WES has been proved as a suitable and determinant technique to find new BC susceptibility candidate genes, as well as variants not previously reported in well-established BC susceptibility genes.

7. Conclusions

1. Although our initial hypothesis was the recessive model of inheritance, in almost all the BC families studied, the model of inheritance that most probably explained the disease was the dominant. This highlights the difficulty of recognizing patterns of inheritance for this disease which is one of the reasons that has hindered the finding of new susceptibility genes through the years. Nonetheless, this does not imply that the recessive model is inexistent, as it still fits our MBC family. It is also possible that a combination of variants fitting the dominant and recessive model may coexist under a polygenic model of inheritance.

2. After the exploration of the recessive model of inheritance, the DNA repairing/maintaining genes, the dominant model of inheritance and case-control association study, 25 variants were found as putative susceptibility alleles for hereditary breast cancer which are candidates for further studies.

3. Finding of clearly pathogenic germline variants in *RECQL5*, which could account for at least 0.42% in BRCAX cases in Spanish population, pointed toward a potential involvement of this DNA helicase in increased breast cancer susceptibility.

4. A truncating mutation in *ATM* gene was identified as the causal variant in a BRCAX family. Mutational screening of the gene revealed that nearly 2% of Spanish BC families are associated to germline pathogenic mutations in the *ATM* gene which strengthens its inclusion in routine genetic tests in clinical setting for BRCAX families to detect *ATM* heterozygous carriers whom could benefit from specific monitoring programs.

5. We have identified variants in seven candidate susceptibility genes for hereditary male breast cancer. From these, the most interesting are *TXNDC5* related with regulation of androgen receptor signaling pathway and *LIMD1* which interacts with the MBC susceptibility gene *BRCA2*, which both require further in-depth analysis.

7. Conclusiones

1. Aunque nuestra hipótesis inicial fue el modelo recesivo de herencia, en casi todas las familias estudiadas, el modelo de herencia más probable para explicar la enfermedad resultó ser el modelo dominante. Esto pone de manifiesto la dificultad para reconocer patrones de herencia específicos en familias con cáncer de mama, siendo esta una de las razones que ha dificultado la identificación de nuevos genes a lo largo de los años. Sin embargo, esto no implica que el modelo recesivo sea inexistente, ya que todavía se ajusta a nuestra familia de cáncer de mama en varón. Por otro lado, es posible que una combinación de variantes que ajustan con el modelo dominante y recesivo coexistan bajo el modelo de herencia poligénica.

2. Después de explorar el modelo recesivo de herencia, los genes de reparación/ mantenimiento del ADN, el modelo dominante de herencia y tras el estudio de asociación de casos y controles, se encontraron 25 variantes candidatas de susceptibilidad a cáncer de mama hereditario que son candidatas a estudios en series más amplias.

3. El hallazgo de mutaciones claramente deletéreas en línea germinal en *RECQL5*, en al menos un 0.42% de las familias BRCAX en población española, señalan hacia un papel potencial de esta DNA helicasa en la susceptibilidad al cáncer de mama.

4. Se identificó una mutación truncante en el gen *ATM* como la variante causal en una familia BRCAX. El análisis mutacional de las regiones codificantes del gen reveló que casi el 2% de las familias españolas con cáncer de mama están asociadas a mutaciones deletéreas en línea germinal en el gen *ATM*, lo cual refuerza su inclusión en pruebas genéticas de rutina en el contexto clínico para familias BRCAX para detectar portadores heterocigotos de mutaciones que podrían beneficiarse de programas de prevención específicos.

5. Hemos identificado variantes en siete genes candidatos de susceptibilidad a cáncer de mama hereditario en varón. De estos, son especialmente interesantes *TXNDC5* asociado con la regulación de la vía de señalización el receptor de andrógenos y *LIMD1* que interacciona con el gen de susceptibilidad para cáncer de mama en varón *BRCA2*, ambos requieren análisis moleculares y estadísticos en profundidad para establecer su relación con la enfermedad.

8. Supplementary Data and Tables

Supplementary Table 1:

Principal function of the genes found through Whole Exome Sequencing

Fam	Gene	Function*	Model of inheritance
1	*DOCK3*	Dedicator Of Cytokinesis 3, family of guanine nucleotide exchange factors (GEFs), modifier of cell adhesion (MOCA) and presenilin-binding protein (PBP), involved in cytoskeletal remodeling and signal transduction.	Recessive
1	*POLQ*	DNA Polymerase Theta, DNA DS break repair, promotes microhomology-mediated end-joining and promotes genomic rearrangements.	DNA repairing gene
1	*CHEK2*	Checkpoint Kinase 2, required for checkpoint-mediated cell cycle arrest, activation of DNA repair and apoptosis in response to the presence of DNA double-strand breaks.	DNA repairing gene
1	*POLG*	DNA Polymerase Gamma, involved in the replication of mitochondrial DNA.	DNA repairing gene
1	*KANK1*	KN Motif And Ankyrin Repeat Domains 1, Cytoeskeleton formation, regulates actin polymerization, tumor suppressor in renal carcinoma, inhibits RhoA activity, cell polarity, among others.	Dominant
1	*CDH15*	Cadherin 15, Encoding calcium-dependent intercellular adhesion glycoproteins, control of morphogenetic processes.	Dominant
1	*ACTR8*	ARP8 Actin-Related Protein 8 Homolog, DNA DS Break Repair and TC-NER, organization of mitotic chromosomes, transcriptional regulation, DNA replication, DNA repair, INO80 complex.	Dominant
1	*CERS5*	Ceramide Synthase 5, Synthesis of ceramide, a lipid molecule that is involved in a several cellular signaling pathways, link with apoptosis, endometrial cancer, upregulated in BC cell lines.	Dominant
1	*KAT5*	Lysine Acetyltransferase 5, MYST family of histone acetyl transferases (HATs), chromatin remodeling, transcription and other nuclear processes by acetylating histone and non histone proteins, DNA repair, Signal transduction, NuA4 complex required for tumor suppression of cell growth, replicative senescence and apoptosis.	Dominant
1	*MTUS1*	Microtubule Associated Tumor Suppressor 1, Cooperates with AGTR2 to inhibit ERK2 activation and cell proliferation. May be required for AGTR2 cell surface expression, inhibits breast cancer cell proliferation, delays the progression of mitosis by prolonging metaphase and reduces tumor growth.	Dominant
1	*PLEC*	Plectin, Interlinking different elements of the cytoskeleton, scaffolding platforms for the assembly, positioning and regulation of signaling complexes, related with EFGR1 signaling, degradation of extracellular matrix.	Dominant
1	*DDX11*	DEAD/H-Box Helicase 11, DNA-dependent ATPase and ATP-dependent DNA helicase that participates in various functions in genomic stability, including DNA replication, DNA repair and heterochromatin organization as well as in ribosomal RNA synthesis.	Dominant
1	*NISCH*	Nischarin, cytoskeletal organization and cell migration by binding to alpha-5-beta-1 integrin.	Dominant
2	*AKR1C3*	Aldo-Keto Reductase Family 1 Member C3, Ovarian steroidogenesis, interconverts active androgens, estrogens and progestins with their cognate inactive metabolites.	Recessive
2	*FBF1*	Fas Binding Factor 1, Keratin-binding protein required for epithelial cell polarization. Involved in apical junction complex (AJC) assembly via its interaction with PARD3. Required for ciliogenesis and Fas induced apoptosis.	Recessive
2	*GLE1*	GLE1 RNA Export Mediator, Export of mature mRNA.	Recessive
2	*KLB*	Klotho Beta protein, is involved in RET signaling, involved in the transcriptional repression of hydroxylase 7-alpha cholesterol, participates and enhances the union of the fibroblast growth factor to its receptors 1 and 4. Participates in signaling pathways from PI3kinases and insulin receptor signaling, its functions affect the metabolism of the glucose, lipids and energy biogenesis.	Recessive
2	*TOPAZ1*	Testis And Ovary Specific PAZ Domain Containing 1, germ cell specific factor, meiotic progression, contribute to silencing of transposable elements and maintenance of genome integrity	Recessive

2	*POLK*	DNA Polymerase Kappa, catalyzes translesion DNA synthesis, which allows DNA replication in the presence of DNA lesions.	DNA repairing gene
2	*RECQL5*	RecQ Like Helicase 5, prevents aberrant homologous recombination by displacing RAD51 from ssDNA, participates in DNA replication, transcription and repair, required for mitotic chromosome separation after cross-over events and cell cycle progress.	DNA repairing gene
	SPRED2	Sprouty Related EVH1 Domain Containing 2, regulate growth factor-induced activation of the MAP kinase cascade, it also has function in stem cell factor receptor binding.	Dominant
2	*DEPDC1B*	DEP Domain Containing 1B, Related with GPCR and signaling by Rho GTPases, related with BC.	Dominant
2	*GPRC5B*	G Protein-Coupled Receptor Class C Group 5 Member B, G-protein coupled receptor activity and G-protein coupled receptor binding.	Dominant
2	*HBP1*	HMG-Box Transcription Factor 1, involved in regulation of C-myc pathway, regulation of wnt mediated β-catenin pathway, regulation of cell cycle, disrupts interaction among DNA and proteins.	Dominant
2	*SYNE1*	Spectrin Repeat Containing Nuclear Envelope Protein 1, May be involved in the maintenance of nuclear organization and structural integrity. Connects nuclei to the cytoskeleton by interacting with the nuclear envelope and with F-actin in the cytoplasm. May be required for centrosome migration to the apical cell surface during early ciliogenesis.	Dominant
2	*AHNAK*	AHNAK Nucleoprotein, blood-brain barrier formation, cell structure and migration, cardiac calcium channel regulation and tumor metastasis.	Dominant
2	*KLF15*	Kruppel Like Factor 15, Transcription factor activity, sequence-specific DNA binding and transcriptional activator activity.	Dominant
2	*HHAT*	Hedgehog acetyltransferase, signaling through GPCR, CDK mediated phosphorylation and removal CDC6, critical role in ER positive, HER2 amplified and hormone resistant breast cancer proliferation.	Dominant
2	*MYBBP1A*	MYB Binding Protein 1a, encodes a nucleolar transcriptional regulator that was first identified by its ability to bind specifically to the Myb proto-oncogene protein, nucleolar stress, tumor suppression and synthesis of ribosomal DNA	Dominant
2	*RANBP10*	RAN Binding Protein 10, Cytoplasmic Guanine nucleotide exchange factor that modulates non-centrosomal microtubules, novel coactivator of androgen receptor.	Dominant
2	*CAPN2*	Calpain 2, calcium-activated neutral proteases, which is a non lysosomal, intracellular cysteine protease, involved in cytoskeletal remodeling and signal transduction.	Dominant
2	*CLSPN*	Claspin, regulator of checkpoint kinase 1 and triggers a checkpoint arrest of the cell cycle in response to replicative stress or DNA damage. The protein is also required for efficient DNA replication during a normal S phase, Adapter protein which binds to BRCA1 and the checkpoint kinase CHEK1 and facilitates the ATR-dependent phosphorylation of both proteins.	Dominant
2	*PTPRG*	Protein Tyrosine Phosphatase, Receptor Type G, molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle and oncogenic transformation.	Dominant
3	*ATM*	Ataxia Telangiectasia Mutated, important cell cycle checkpoint kinase that phosphorylates; thus, it functions as a regulator of a wide variety of downstream proteins, including tumor suppressor proteins p53 and BRCA1, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1, required for cell response to DNA damage and for genome stability.	DNA repairing gene
4	*LIMD1*	LIM Domains Containing 1, Adapter or scaffold protein which participates in the assembly of numerous protein complexes and is involved in several cellular processes such as cell fate determination, cytoskeletal organization, repression of gene transcription, cell-cell adhesion, cell differentiation, proliferation and migration.	Recessive
4	*TXNDC5*	Thioredoxin Domain Containing 5, encodes a protein-disulfide isomerase, reduces insulin disulfide bonds, participates in the modulation of androgen receptor signaling pathway.	Recessive
4	*PLCB2*	Phospholipase C Beta 2, production of the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) is mediated by activated phosphatidylinositol-specific phospholipase C enzymes. Important for cell signaling pathways mediators of survival and proliferation.	Recessive
4	*GATM*	Glycine Amidinotransferase, mitochondrial enzyme involved in creatine biosynthesis, whereby it catalyzes the transfer of a guanido group from L-arginine to glycine, resulting in guanidinoacetic acid, the immediate precursor of creatine, plays a vital role in energy, aberrant chimeric fusion alter cell migration and invasion.	Recessive

4	*PTPRH*	Protein Tyrosine Phosphatase, Receptor Type H, involved in cellular processes including cell growth, differentiation, mitotic cycle and oncogenic transformation.	Recessive
4	*YLPM1*	YLP Motif Containing 1, Plays a role in the reduction of telomerase activity during differentiation of embryonic stem cells.	Recessive
4	*TKTL1*	Transketolase Like 1, transketolase that acts as a homodimer and catalyzes the conversion of sedoheptulose 7-phosphate and D-glyceraldehyde 3-phosphate to D-ribose 5-phosphate and D-xylulose 5-phosphate, links the pentose phosphate pathway with the glycolytic pathway.	X-linked

*Information retrieved from http://www.genecards.org/, last accessed June 2017

Supplementary Table 2:

Primers designed for Sanger sequencing validation of BC candidate susceptibility alleles

Oligo Name	Tm (ºC)	GC %	Sequence (5'-3')	Product length	Note
CLIP4_E11F	64	42.8	AAGATTTGCCTTGACAAGCTG	205	
CLIP4_E11F	64	45	ATGAAAGGGAAGGGCTTTTG		
YLPM1_E5F	64	47.6	CAGGATTGGTCAAGCAAGAAG	230	
YLPM1_E5F	64	60	AAGAGGTCCCCTCTCCTGAC		
TOPAZ1_Q483_E2F	64	37.5	TGTCCCAGAAACAGTAGAAAAAGA	342	
TOPAZ1_Q483_E2R	66	45	GAAGCCTTTTTCCAACACCA		
TOPAZ1_K796_2EF	64	40.9	GCAAGGTAACAATAGCAAACCA	175	
TOPAZ1_K796_2ER	64	50	GGACACTGGTCGGAATGTTT		
GLE1_C39_2EF	64	45.8	GATGTTCAGTAAGGCATGTAGTGG	167	
GLE1_C39_2ER	64	55	CTCCTGCATATGGGGTAGGA		
TNN_E2_F	64	50	AAGATCGATGTGCCCAAGTC	225	
TNN_E2_R	64	55	CTCCACCATCTCTTCCTCCA		
SUPT7L_E5_F	64	60	CTGAGAAGGCCACAGAGGAC	195	
SUPT7L_E5_R	64	52.6	TTGAAGCCTGTGGTGAAGC		
SUPT7L_E3_F	64	55	TCAGCAGCAGACAGAAGGTG	252	
SUPT7L_E3_R	64	40	CCTCGCATGAAAATGGAAAT		
BIRC6_E10_F	62	45	CATTTTGCTTCCACCCGATA	241	
BIRC6_E10_R	62	55	TGTGTCCTGTTCCTCAGTGC		
LIMD1_E1_F	62	57.8	CTTTGGTCCACTGCCTCCT	210	
LIMD1_E1_R	62	45	CACATTGGGTTTGGACATCA		
PLCB2_E9_F	62	50	CATGCTAAGGCCAAACCCTA	153	
PLCB2_E9_R	62	55	GATGCCACTGGGCTCATACT		
GATM_E2_F	62	55	TCCATCTCCACTTCCTCCTC	228	
GATM_E2_R	62	50	GGGTCCCATTCGTTGTAAGA		
HAS1_E5F	64	55	ACCATGCGTGGATGACCTAC	230	
HAS1_E5R	64	55	CAGGAGGCCACACATGTAGA		
PTPRH_E11F	64	55	AGGGTTGGAGACAGATGAGC	210	
PTPRH_E11R	62	57.8	AGCGGTCCTCCTCCTTCTT		
MXRA5_E5F	62	50	AAAGTTGCTCCGTCATCCAC	218	
MXRA5_E5R	62	55	TCCACCCAAGCTGTAGGAAC		
MAMLD1_E3F	64	55	CACTGAGCTGGTGCAGGTAA	208	
MAMLD1_E3R	64	55	GCCTCAGTTTCCCCACTGTA		
TKTL1_E10F	64	45	AACCATGGCTCCATTTATGC	217	
TKTL1_E10R	62	40.9	GCAAAGAAAGGATGATGTCTCA		
TXNDC5_E10F	62	45	GCAATTTGCATGTGATCTGG	284	
TXNDC5_E10R	62	50	TTTCTGAACAGCCACCACTG		
CLSPN_E5F	64	50	GTGGGGTCCATTCATTTGAG	220	
CLSPN_E5R	64	34.7	GAGAATTTGTTCAATCCAATCCA		
NISCH_E16F	64	57.8	CATCCTCAAGGTGCTGTGG	237	
NISCH_E16R	62	55	CGTACCTTGAGCACCAGACA		

DDX11_E4F	62	55	GGTGCAGCTCAAGTATGCAG	232	For splicing variant
DDX11_E7R	62	55	GGCTCTTCTTCACCTCATGC		
AKR1C3_E2F	64	50	TACCTTTGGTTGCTCCTCCA	214	
AKR1C3_E2R	64	43.4	CACAAGCTCATCATAGACACAGT		
DOCK3_E9F	64	36	AATCTCCTTCAGTAGTGTGAAAGA	296	
DOCK3_E9R	62	42.8	AGAACCTCACCAGAATTTCCA		
DOCK3_E53F	62	50	TCTAACATGCCCACCCTTTC	298	
DOCK3_E53R	62	61.1	CTGGGCCTTGATGGAGTC		
POLG_E16F	64	55	TCTGCTGAGTGGTTGTAGGG	245	
POLG_E16R	64	60	GGCCAGAGGTACAGAGGTCA		
FBF11_E19F	64	61.1	CAGCCCTCTGCCTTCCTT	299	
FBF11_E19R	62	55	TACTCCTTGCAAGCCTGCTC		
FBF11_E18F	62	63.1	CGTGGAGTACAGCCACCTC	206	
FBF11_E18R	62	55	TTGGGGTCAGTGTGAGATCC		
FBF11_E11F	64	60	GAGCCCCTCATACCTACAGG	249	
FBF11_E11R	64	45	AGGTTGTGATGAAAGCCTGA		
CHEK2_E16F	64	50	CAAATGCCCCCACTTTACTG	458	
CHEK2_E16R	64	55.5	CTGCACCACTGCACTTCA		
ATM_eE36_NF	62	30.4	TTTGACAAAGAAAACCCTTTTGA	166	LOH in tumor
ATM_e36_NR	62	34.7	TTCTTACTTCACACATTGGCTTT		

Supplementary Table 3:

Primers designed for RECQL5 Sanger sequencing validation of candidate variants

Oligo Name	Tm (ºC)	GC %	Sequence (5'-3')	Product length	Note
RECQL5_EXON3F	64	50	AAAGGCAGGAAGCACTGAGA	216	
RECQL5_EXON3R	64	45	ACCTTCCCCGAATCTTTTTG		
RECQL5_EXON4F	64	55	GGTCCAAGAGGACGTGTTTG	334	
RECQL5_EXON4R	64	50	GCCAGGAAACCAAAGAGACA		
RECQL5_EXON9F	64	57,8	GATTCAGGCTCGGGTGAAC	312	
RECQL5_EXON9R	66	66,6	GGCTTCTCGGCCCCTTAC		
RECQL5_EXON17F	64	55	GATGCCTGCTCTTCCCTCTA	292	
RECQL5_EXON17R	64	60	CTCCCCTCCTCCAGAAGACT		
RECQL5_EXON6F	64	50	TCTCCATCTTCCCCTCCTTT	225	
RECQL5_EXON6R	64	52,6	ACATGCACCTGGTCCTTTG		
RECQL5_EXON20F	64	55	GCTAACGCTCACCTCCTCAT	280	
RECQL5_EXON20R	64	50	GTCATCCCCAAAGCCAAGTA		
RECQL5_EXON15F2	64	63,1	CACAGCCTCCCTGCCTATC	337	
RECQL5_EXON15R2	64	55	GCTGGCTGGAGCAGTAACTC		
RECQL5_EXON10F	64	50	CTGGTAGGACATCAGGGACATT	357	
RECQL5_EXON10R	64	45	TCTTCTATCTTGGGGTCTTTGC		
RECQL5_EXON15F	62	57,8	AGTGGTCAGTTGCCTGTGG	315	
RECQL5_EXON15R	62	50	AGTGAAGCCTTTCCTGGTCA		
RECQL5_INTRON19F	62	57,1	CCACCTCTCACACTTGCTGAC	190	
RECQL5_INTRON19R	62	50	CCATGGAAGAAGTGCCTGAT		
RECQL5_INTRON11F	62	50	AATGTTCCTGGAGGCTTGTG	198	
RECQL5_INTRON11R	62	55	TGCCCGTCTCTTACCTTCAC		
RECQL5_EXON14F	62	55	CAAGAGCTCAGCAGGGAGAA	249	
RECQL5_EXON14R	62	66,6	CCTGAGCCCAGAGAGCTG		
RECQL5_EXON5F	64	55	AGCAGGGTGCATTACACTGG	226	
RECQL5_EXON5R	64	50	AAGAAGCCTCTGAGGGTGAA		
RECQL5_EXON2F	64	50	GCAAAACAGAGGGTTCTCGT	295	
RECQL5_EXON2R	64	60	GAGGCAGTACGAAGGGTGAG		
RECQL5_EXON16F	64	45	CCCTGAGAAGAAGGCAAAAA	353	
RECQL5_EXON16R	64	55	GCCGTGTAGGTTCCAGAAGA		
RECQL5_EXON13F	64	60	GTAAGAGACGGGCAGCTCCT	234	
RECQL5_EXON13R	64	60	GTTGACAGGGTCCTGCAGTC		
RECQL5_EXON19F	64	60	CTTCTCCCTCTCCCTTCCAG	169	
RECQL5_EXON19R	64	55	ATGAGGAGGTGAGCGTTAGC		
RECQL5_E1F	62	50	CAGCCAGCAGCCTTTAATTC	376	For splicing variant
RECQL5_E4R	62	50	TGCAGAGAGCTTCGAGTTCA		
RECQL5_E8F	62	57,8	GATGCCCTGGTGACCTTCT	299	For splicing variant
RECQL5_E10R	62	55	GCCTCATCTCTGCCTTCATC		
RECQL5_EXON3F	64	55	AGTCCGGAGTACGCTGAAGA	351	For splicing variant
RECQL5_EXON3R	64	50	AGCTTCATCCACCACCAAGT		

RECQL5_EXON17F	64	55	GATGGCCCCAGAGAAGTACA	320	For splicing variant
RECQL5_EXON17R	64	45	ACAACATTTGCAGCCTCCTT		
RECQL5_EXON16F2	64	55	AGCTGGTTTCCCCAAAGG	317	For splicing variant
RECQL5_EXON16R	64	55	GCCGTGTAGGTTCCAGAAGA		

Supplementary Table 4:

Candidate variants for hereditary breast cancer discarded for further studies based on Case-Control association studies

Family	Gene	Ch	Chromo-somal position	Variant	Protein effect	Functional Class	ID	Model of inheritance	MAF ExAc (Total)[h]	Nº Homozygotes ExAc (European)	Eliminated due to:
1	*DOCK3*	3	51127722	c.653T>C	F218S	Missense	NR	Recessive	NR	NR	a
1	*DOCK3*	3	51418611	c.5714C>T	S1905L	Missense	rs201507848	Recessive	0.0003732	0/120582	c
1	*CHEK2*	22	29083961	c.1685G>T	R562L	Missense	NR	DNA repairing systems	0.0002315	NA	c
1	*POLG*	15	89865073	c.2492A>G	Y831C	Missense	rs41549716	DNA repairing systems	0.006277	NA	c
1	*DDX11*	12	31238059	c.637A>G	R213G	Missense	rs2536756	Dominant	0.002273	NA	a
1	*NISCH*	3	52521676	c.2168G>A	R723H	Missense	NR	Dominant	0.0000165	NA	d
2	*AKR1C3*	10	5138747	c.230A>G	E77G	Missense	rs11551177	Recessive	0.05018	212/120836	a,b
2	*FBF1*	17	73915803	c.2039C>T	A680V	Missense	rs113062332	Recessive	0.02305	45/109134	c
2	*TOPAZ*	3	44286015	c.2017C>G	P673A	Missense	rs17646517	Recessive	0.01783	9/20860	c
2	*TOPAZ*	3	44373498	c.5074C>G	H1692D	Missense	rs533942526	Recessive	0.00009226	0/21678	e
2	*KLB*	4	39448672	c.2329_2331del	F777del	Frameshift	NR	Recessive	0.002272	NR	e
2	*KLB*	4	39450295	c.3124G>A	V1402I	Missense	rs143809363	Recessive	0.003605	5/119004	e
2	*POLK*	5	74892973	c.2455A>C	N819H	Missense	rs185752953	DNA Repairing systems	0.001024	NA	d
2	*SPRED2*	2	65561832	c.280G>A	D94N	Missense	NR	Dominant	0.000008237	NA	a
2	*MYBBP1A*	17	4455198	c.1000G>A	G334R	Missense	rs138633396	Dominant	0.001129	NA	c
2	*CAPN2*	1	223900358	c.16G>A	A6T	Missense	rs375899944	Dominant	0.0001054	NA	g
2	*PTPRG*	3	62189151	c.1682C>T	S561Y	Missense	rs201820508	Dominant	0.00043	NA	g
2	*CLSPN*	1	36228012	c.815C>T	T272M	Missense	rs141350492	Dominant	0.0001566	NA	c*

NR, No report, NA Not applicable under the model of inheritance studied
a No results in the Openarray
b No results in the Openarray and Taqman probe
c High frequency both in cases and controls of the Spanish cohort (*) equal frequency in cases and controls
d Did not fit with the model of inheritance studied and/or higher recurrence in controls against cases

e Due to technical constraints, were not included in OpenArray. They will be included in other studies
g Since the beginning, it was discarded from the Openarray because the design of the probe did not worked either in the CEGEN facility or in the fabric where it was designed.
h Total values for EXAC include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations.

Supplementary Table 5:

Rare[1] variants detected during the screening of *ATM* in analyzed in 392 Spanish BRCAX families by NGS

Nucleotide change[2]	Protein change	rs number	ClinVar[3]	MAF	SIFT Prediction	Polyphen Prediction
c.1229T>C	p.Val410Ala	rs56128736	Conflicting interpretations of pathogenicity: Benign (2), Likely benign (1), Uncertain significance (2)	0,003277	deleterious	benign
c.1464G>T	p.Trp488Cys	rs377597949	Clinical significance: Uncertain significance	4,50E-05	tolerated	benign
c.1595G>A	p.Cys532Tyr	rs35963548	Clinical significance: Uncertain significance	0,0001351	deleterious	benign
c.1744T>C	p.Phe582Leu	rs2235006	Clinical interpretation: Benign/Likely benign, not provided	0,001004	tolerated	benign
c.1810C>T	p.Pro604Ser	rs2227922	Conflicting interpretations of pathogenicity: Benign (2), Likely benign (2), Uncertain significance (1)	0,001	tolerated	possibly_damaging
c.1966A>G	p.Thr656Ala	Not reported	Not reported	Not reported	deleterious	benign
c.2289T>A	p.Phe763Leu	rs34231402	Conflicting interpretations of pathogenicity: Likely benign (1), Uncertain significance (2)	0,0004048	tolerated	benign
c.2494C>T	p.Arg832Cys	rs2229022	Clinical significance: Uncertain significance	0,0002414	tolerated	benign
c.2611G>C	p.Glu871Gln	Not reported	Not reported	Not reported	tolerated	benign
c.2921+19dup	p.?	rs575967175	Clinical significance: Benign	Not reported	–	–
c.3403-15T>A	p.?	rs79701258	Clinical significance: Benign	0,002607	_	_
c.4060C>A	p.Pro1354Thr	rs145119475	Clinical significance: Uncertain significance	0,0003757	tolerated	benign
c.4388T>G	p.Phe1463Cys	rs138327406	Conflicting interpretations of pathogenicity: Benign (2), Likely benign (1), Uncertain significance (2)	0,002276	deleterious	probably damaging
c.4523A>G	p.Tyr1508Cys	Not reported	Not reported	Not reported	tolerated	benign
c.5071A>C	p.Ser1691Arg	rs1800059	Clinical significance: Benign/Likely benign, not provided	0,003132	tolerated	benign
c.5497-15G>C	p.?	rs3092828	Benign	0,004346	_	_
c.5558A>T	p.Asp1853Val	rs1801673	Clinical significance: Benign/Likely benign, not provided	0,006934	deleterious	benign
c.5611A>C	p.Thr1871Pro	Not reported	Not reported	Not reported	tolerated	benign
c.6067G>A	p.Gly2023Arg	rs11212587	Conflicting interpretations of pathogenicity: Benign (1), Likely benign (2), Uncertain significance (1)	0,002324	deleterious	probably_damaging
c.7187C>G	p.Thr2396Ser	rs370559102	Clinical significance: Uncertain significance	0,0002555	tolerated	benign
c.7788+8G>T	p.?	rs112775908	Clinical significance: Benign/Likely benign	0,002599	_	_
c.79G>A	p.Val27Ile	rs754770960	Clinical significance: Uncertain significance	1,50E-05	tolerated	benign

c.8113G>A	p.Val2705Ile	rs587779870	Clinical significance: Uncertain significance	1,50E-05	tolerated	benign
c.8584+19T>C	p.?	rs772128061	Not reported	1,50E-05	_	_
c.931A>G	p.Ile311Val	Not reported	Clinical significance: Uncertain significance	Not reported	tolerated	benign
c.984_986del	p.Ser328del	Not reported	Clinical significance: Uncertain significance	Not reported	tolerated	benign
c.998C>T	p.Ser333Phe	rs28904919	Clinical significance: Benign/Likely benign, not provided	0,001398	deleterious	possibly_damaging

[1] Variants were considered as rare and included in this table when MAF (Minor Allele Frequency) reported in ExAc (http://exac.broadinstitute.org) for European (non-finish) population was lower than 0.01
[2] GenBank reference sequence NM_000051.3 with numbering starting at the A of the first ATG, following the HGVS guidelines, www.hgvs.org/mutnomen
[3] Summary of classification as appearing in ClinVar database. None of the variants in this table were classfied in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) with three asterisks and were therefore considered as Unknown Significance Variants.

Supplementary Table 6:

Candidate deleterious mutations found in other 20 susceptibility genes or candidate susceptibility genes analyzed in 392 Spanish BRCAX families by NGS

Gene[2]	Nucleotide change	Protein change	GenBank refseq	rs number	ClinVar[3]	MAF[4]	Nº of carriers in 392 BRCAX families
BARD1	c.334C>T	p.Arg112*	NM_000465.3	rs758972589	Clinical significance:Pathogenic**	0	1
CHEK2	c.507del	p.Asp169Glufs*5	NM_001005735.1	Not reported	Not reported	Not reported	1
CHEK2	c.1229del	p.Thr410Metfs*15	NM_001005735.1	rs555607708	Clinical significance:Pathogenic**	0,002341	1
FANCE	c.1239dupT	p.Pro414Serfs*54	NM_021922.2	Not reported	Not reported	Not reported	1
FANCI	c.3626_3627delGT	p.Cys1209Leufs*10	NM_001113378.1	rs770318990	Clinical significance:Pathogenic*	0,00E+00	1
FANCL[5]	c.1096_1099dup	p.Cys367Leufs*3	NM_001114636.1	Not reported	Not reported	Not reported	1
FANCM	c.855del	p.Val286Leufs*30	NM_020937.2	Not reported	Not reported	Not reported	2
FANCM[6]	c.1511_1515del	p.Arg504Asnfs*29	NM_020937.2	Not reported	Not reported	Not reported	1
FANCM	c.1972C>T	p.Arg658*	NM_020937.2	rs368728266	Not reported	Not reported	2
FANCM[7]	c.5791C>T	p.Arg1931*	NM_020937.2	rs144567652	Not reported	0,0009456	1
PALB2	c.172_175del	p.Gln60Argfs*7	NM_024675.3	rs180177143	Conflicting interpretations of pathogenicity: Likely pathogenic(1);Pathogenic(4);Uncertain significance(1)	Not reported	1
PALB2	c.757_758del	p.Leu253Ilefs*3	NM_024675.3	rs180177092	Clinical significance:Pathogenic**	1,50E-05	1
PALB2	c.1188C>A	p.Cys396*	NM_024675.3	Not reported	Not reported	Not reported	1
PALB2	c.2257C>T	p.Arg753*	NM_024675.3	rs180177110	Clinical significance:Pathogenic**	4,50E-05	1
PALB2	c.1653T>A	p.Tyr551*	NM_024675.3	rs118203997	Clinical significance:Pathogenic**	Not reported	1
RAD50	c.2517dup	p.Asp840Argfs*5	NM_005732.3	Not reported	Clinical significance:Pathogenic*	Not reported	1
RAD51C	c.709C>T	p.Arg237*	NM_058216.1	rs770637624	Clinical significance:Pathogenic**	1,50E-05	1
RAD51D	c.94_95del	p.Val32Phefs*38	NM_002878.3	Not reported	Clinical significance:Pathogenic*	Not reported	1
SLX4	c.4739+1G>T	p.?	NM_032444.2	rs759186986	Not reported	1,50E-05	1
SLX4	c.5154-1G>T	p.?	NM_032444.2	Not reported	Not reported	Not reported	1
XRCC2	c.96del	p.Phe32Leufs*30	NM_005431.1	rs730882048	Conflicting interpretations of pathogenicity: Pathogenic(1);Uncertain significance(1)	1,36E-04	1

[1] Only mutations giving rise to a premature stop codon or affecting consensus splicing sites (+/-1 and +/- 2) and/or reported in ClinVar as pathogenic with three asterisks*** were considered
[2] The genes included in the analysis are detailed in the methods section
[3] Summary of classification as appearing in ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/)
[4] MAF (Minor Allele Frequency) reported in ExAc (http://exac.broadinstitute.org) for European (non-finish) population
[5] The patient carrying this mutation also carried the mutation c.855del in *FANCM*
[6] The patient carrying this mutation also carried the mutation c.4739+1G>T in *SLX4*
[7] This mutation was found in homozigosity

Supplementary Table 7:

Candidate variants for hereditary male breast cancer discarded from initial studies

Gene	Ch	Chromo-somal position	Variant	Protein effect	Functional Class	ID	Model of inheritance	MAF ExAc (Total)[a]	Nº Homozygotes ExAc (European)	Eliminated due to:
TNN	1	175046826	c.272G>A	R91H	Missense	rs41266078	Recessive	0,02902	65/121020	[c]
BIRC6	2	32641040	c.2681C>T	T894M	Missense	rs34996177	Recessive	0.02384	54/121272	[d]
SUPT7L	2	27883899	c.371A>C	N124T	Missense	rs147739348	Recessive	0,001506	2/120148	[d]
SUPT7L	2	27878328	c.886G>A	A296T	Missense	rs201935160	Recessive	0,0004817	0/120398	[d]
GPR113	2	26534250	c.2346A>G	I782M	Missense	rs114354727	Recessive	0,03048	90/121242	[e]
MGAM	7	141765575	c.4714A>T	M1572L	Missense	rs4507684	Recessive	0,03956	621/116030	[e]
HAS1	19	52217128	c.1289C>T	A430V	Missense	rs34682338	Recessive	0.008392	6/108680	[d]
MXRA5	X	3240238	c.3488G>A	R1163H	Missense	rs139106444	X-linked	0,005675	164//82817 [b]	[d]
MAMLD1	X	149631097	c.156A>G	P52P	Silent	rs16996606	X-linked	0,005213	90/83249 [b]	[d]

a Total values for EXAC include European (non Finish and Finnish), Latino, South Asian, African and East-Asian populations.
b Refers to the number of hemyzogotes
c False positive
d Did not fit the model of inheritance when the segregation analysis were extended to other members of the family
e Presented a high frequency in EXAC population and/or there were technical constraints to consider the variant for further analysis

9. References

Ali AMG, Schmidt MK. 2014. Alcohol Consumption and Survival after a Breast Cancer Diagnosis: A Literature-Based Meta-analysis and Collaborative Analysis of Data for 29,239 Cases. Cancer Epidemiol Biomarkers Prev. 23(6): 934–945

Aloraifi F, Boland MR et al., 2015. Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. Surg Oncol. 24(2): 100-9

American Cancer Society. Breast Cancer, Facts and Figures. 2015-2016. Atlanta: American Cancer Society. Inc. 2015

Antoniou AC, Easton DF. 2003. Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol. 25 (3): 190-202

Antoniou AC, Easton DF. 2006. Models of genetic susceptibility to breast cancer. Oncogene. 25 (43): 5898-905

Antoniou AC, Casadei S et al., 2014. Breast-cancer risk in families with mutations in PALB2. N Engl J Med 371 (6): 497-506

Apostolou P, Fostira F. 2013. Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int. 2013: 747318

Armstrong L, Lako M et al., 2004. A role for nucleoprotein Zap3 in the reduction of telomerase activity during embryonic stem cell differentiation. Mech Dev. 121(12):1509-22

Aygün O, Xu X et al., 2009. Direct inhibition of RNA polymerase II transcription by RECQL5. J Biol Chem. 284 (35): 23197-203

Aygün O, Svejstrup JQ. 2010. RECQL5 helicase: connections to DNA recombination and RNA polymerase II transcription.DNA Repair (Amst) 9(3): 345-53

Azim HA, Partridge AH. 2014. Biology of breast cancer in young women. Breast Cancer Res. 16: 427

Bahcall O, 2013. Common variation and heritability estimates for breast, ovarian and prostate cancers. Nat Genet. Access only available: doi: 10.1038/ngicogs.1

Bai SW, Herrera-Abreu MT et al., 2011. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. BMC Biol. 9:54

Bao RB, Huang L et al., 2014. Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. Cancer Inform. 13(Suppl 2):67-82

Baumann C, Behbahani TE. 2012. EAU guidelines for prostate cancer: to screen or not to screen? Actas Urol Esp. 36 (1):1

Bassi C, Li YT et al., 2016. The acetyltransferase Tip60 contributes to mammary tumorigenesis by modulating DNA repair. Cell Death Differ. 23(7):1198-208

Beggs AD, Hodgson SV. 2009. Genomics and breast cancer: the different levels of inherited susceptibility. Eur J Hum Genet. 17 (7): 855-6

Bertagnolo V, Benedusi M et al., 2006. PLC-beta2 is highly expressed in breast cancer and is associated with a poor outcome: a study on tissue microarrays. Int J Oncol. 28(4):863-72

Bertagnolo V, Benedusi M et al., 2007. Phospholipase C-beta 2 promotes mitosis and migration of human breast cancer-derived cells. Carcinogenesis. 28(8):1638-45

Bertier G, Hétu M et al., 2016. Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. BMC Med Genomics. 9(1):52

Bohr VA. 2008. Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. Trends Biochem Sci. 33(12): 609-20

Boyd NF, Lockwood GA et al., 1998. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 7(12):1133-44

Bradbury AR, Olopade OI. 2007. Genetic susceptibility to breast cancer. Rev Endocr Metab Disord. 8(3): 255-67

Broeks A, Urbanus JH et al., 2000. ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. Am J Hum Genet 66 (2):494-500

Broeks A, Schmidt MK et al., 2011. Low penetrance breast cancer susceptibility loci are associated withs pecific breast tumor subtypes: findings from the Breasy Cancer Association Consortium. Hum Mol Genet. 20(16): 3289-303

Brunet J, Gutierrez-Enriquez S et al., 2008. ATM germline mutations in Spanish early-onset breast cancer patients negative for BRCA1/BRCA2 mutations. Clin Genet 73 (5):465-73

Bush WS, Haines J. 2010. Overview of linkage analysis in complex traits. Curr Protoc Hum Genet. Chapter 1. Unit 1.9.1-18

Byrd PJ, McConville CM et al., 1996. Mutations revealed by sequencing the 5' half of the gene for ataxia telangiectasia. Hum Mol Genet 5 (1):145-9

Cancer Genome Atlas Network et al., 2012. Comprehensive molecular portraits of human breast tumours. Nature. 490(7418):61-70

Cavalieri S, Funaro A et al., 2006. ATM mutations in Italian families with ataxia telangiectasia include two distinct large genomic deletions. Hum Mutat 27 (10):1061

Ceccaldi R, Liu JC et al., 2015. Homologous-recombination-deficient tumors are dependent on Polθ-mediated repair. Nature. 518(7538):258-62

Chong JX, Buckingham KJ et al., 2015. The genetic basis of mendelian phenotypes: discoveries, challenges and opportunities. Am J Human Genet. 97(2): 199-215

Claes K, Depuydt J et al., 2013. Variant ataxia telangiectasia: clinical and molecular findings and evaluation of radiosensitive phenotypes in a patient and relatives. Neuromolecular Med 15 (3):447-57

Collins A, Politopoulos I. 2011. The genetics of breast cancer: risk factor for disease. Appl Clin Genet. 4:11-19

COMPLEXO, Southey MC et al., 2013. COMPLEXO: identifying the missing heritability of breast cancer via next generation collaboration. Breast Cancer Res. 15(3): 402

Cool M, Jolicoeur P. 1999. Elevated frequency of loss of heterozigosity in mammary tumors arising in mouse mammary tumor virus/neu transgenic mice. Cancer Res. 59(10):2438-44

Croteau DL, Popuri V et al., 2014. Human RecQ helicases in DNA repair, recombination, and replication. Annu Rev Biochem. 83:519-52

Cui J, Antoniou AC et al., 2001. After BRCA1 and BRCA2-what's next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. Am J Hum Genet. 68 (2): 420-31

Cybulski C, Carrot-Zhang J et al., 2015. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet 47 (6):643-6

Dai CH, Chen P et al., 2016. Co-inhibition of polθ and HR genes efficiently synergize with cisplatin to suppress cisplatin-resistant lung cancer cells survival. Oncotarget. 7(40):65157-70

Das Thakur M, Feng Y et al., 2010. Ajuba LIM proteins are negative regulators of the Hippo signaling pathway. Curr Biol. 20(7):657-62

Dawn Teare M, Barrett JH. 2005. Genetic linkage studies. Lancet. 366 (9490): 1036-44

Di Lauro L, Barba M et al., 2015. Androgen receptor and antiandrogen therapy in male breast cancer. Cancer Lett. 368(1):20-5

Di Marco S, Hasanova Z et al., 2017. RECQ5 helicase cooperates with MUS81 endonuclease in processing stalled replication forks at common fragile sites during mitosis. Mol cell. 66(5):658-71

Diaz-Moralli S, Aguilar E et al., 2016. A key role for transketolase-like 1 in tumor metabolic reprogramming. Oncotarget. 7(32):51875-97

Diez O, Osorio A et al., 2003. Analysis of BRCA1 and BRCA2 genes in Spanish breast/ovarian cancer patients: a high proportion of mutations unique to Spain and evidence of founder effects. Hum Mutat 22 (4):301-12

Dupont WD, Parl FF et al., 1993. Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer. 71(4): 1258- 65

Easton DF, Pooley KA et al., 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 447(7148):1087-93

Easton DF, Leseur F et al., 2016. No evidence that protein truncating variants in *BRIP1* are associated with breast cancer risk: implications for gene panel testing. J Med Genet. 53(5):298-309

Ellis NA, Groden J et al., 1995. The Bloom's syndrome gene product is homologous to RecQ helicases. Cell. 83(4) :655-66

Escamilla-Powers JR, Daniel CJ et al., 2010. The tumor suppressor protein HBP1 is a novel c-myc-binding protein that negatively regulates c-myc transcriptional activity. J Biol Chem. 285(7):4847-58

Fachal L, Dunning AM. 2015. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr Opin Genet Dev. 30:32-41

Farmer H, McCabe N et al., 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.Nature. 434(7035):917-21.

Feng BJ, Tavtigian SV et al., 2011. Design considerations for massively parallel sequencing studies of complex human disease. PLoS One. 6(8):e23221

Ferlay J, Soerjomataram I et al., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 136 (5):E359-86

Fischhaber PL, Gerlach VL et al., 2002. Human DNA polymerase kappa bypasses and extends beyond thymine glycols during translesion synthesis in vitro, preferentially incorporating correct nucleotides. J Biol Chem. 277(40):37604-11

FitzGerald MG, Marsh DJ et al., 1998. Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. Oncogene. 17 (6): 727-31

Földi M, Stickeler E et al., 2007. Transketolase protein TKTL1 overexpression: A potential biomarker and therapeutic target in breast cancer. Oncol Rep. 17(4):841-5

Foulkes WD. 2008. Inherited susceptibility to common cancers. N Engl J Med. 359 (20): 2143-53

Foulkes WD, Shuen AY. 2013. In brief: BRCA1 and BRCA2. J Pathol. 230(4):347-9

Frank B, Bermejo JL. 2007. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. Carcinogenesis. 28(7):1442-5

Freisinger F, Domchek SM. 2008. Clinical implications of low-penetrance breast cancer susceptibility alleles. Curr Oncol Rep. 11 (1): 8-14

Garcia MJ, Benitez J. 2008. The Fanconi anemia/BRCA pathway and cancer susceptibility. Searching for new therapeutic targets. Clin Transl Oncol. 10(2): 78-84

Giordano SH, Cohen DS et al., 2004. Breast carcinoma in men: a population-based study. Cancer 101(1):51-7

Goldhirsch A, Wood WC et al., 2013. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO 24, 2206-23

Gonzalez KD, Noltner KA et al., 2009. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. J Clin Oncol. 27(8):1250-6

Goto M, Miller RW et al., 1996. Excess of rare cancers in Werner syndrome (adult progeria). Cancer Epidemiol Biomearkers Prev. 5(4): 239-46

Goullet de Rugy T, Bashkurov M et al., 2016. Excess Polθ functions in response to replicative stress in homologous recombination-proficient cancer cells. Biol Open. 5(10):1485-1492

Gracia-Aznarez FJ, Fernandez V et al., 2013. Whole exome sequencing suggests much of non BRCA1/2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. PLoS One. 8 (2):e55681

Grada A, Weinbrecht K. 2013. Next-generation sequencing: methodology and application. J Invest Dermatol. 133(8):e11

Gradishar WJ, Anderson BO et al., 2017. NCCN Guidelines® insights breast cancer, version 1.2017: Featured updates to the NCCN Guidelines. JNCCN J National Comprehensive Cancer Network. 15(4): 433-51

Graffeo R, Livraghi L et al., 2017. Time to incorporate germline multigene panel testing into breast and ovarian cancer patient care. Breast Cancer Res Treat. 160(3):393-410

Graña B, Fachal L et al., 2011. Germline ATM mutational analysis in BRCA1/BRCA2 negative hereditary breast cancer families by MALDI-TOF mass spectrometry. Breast Cancer Res Treat. 128(2):573-9

Gu Y, Liu S et al., 2017. Oncogenic miR-19a and miR-19b co-regulate tumor suppressor MTUS1 to promote cell proliferation and migration in lung cancer. Protein Cell. 8(6):455-66

Hankin JH, Rawlings V. 1978. Diet and breast cancer: a review. Am J Clin Nutr. 31 (11): 2005-16

Hearle N, Schumacher V et al., 2006. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res. 12(10): 3209-15

Heikkinen K, Rapakko K et al 2006. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis. 27(8): 1593-9

Herceg Z, Wang ZQ. 2001. Functions of poly(ADP-ribose) polymerase (PARP) in DNA repair, genomic integrity and cell death. Mutat Res. 477(1-2):97-110

Hilbers FS, Meijers CM et al., 2013. Exome sequencing of germline DNA from Non BRCA1/2 familial breast cancer cases selected on the basis of aCGH tumor profiling. PLoS One. 8 (1):355734

Hirotsu Y, Nakagomi H et al., 2015. Multigene panel analysis identified germline mutations of DNA repair genes in breast and ovarian cancer. Mol Genet Genomic Med 3 (5):459-66

Hönigchmid P. 2012. Improvement of DNA- and RNA-Protein Binding Prediction. PhD thesis. Munich, Germany.

Horna-Terrón E, Pradilla-Dieste A et al.,2014. TXNDC5, a newly discovered disulfide isomerase with a key role in cell physiology and pathology. Int J Mol Sci. 15(12):23501-18

Hou X, Li T et al., 2016. Novel BRCA2-Interacting Protein, LIMD1, Is Essential for the Centrosome Localization of BRCA2 in Esophageal Cancer Cell. Oncol Res. 24(4):247-53

Hu Y, Raynard S et al., 2007. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. Genes Dev. 21 (23): 3073-84

Hu Y, Lu X et al., 2010. Effect of Recql5 deficiency on the intestinal tumor susceptibility of Apc (min) mice. World J Gastroenterol. 16(12): 1482-6

Islam MN, Fox D 3rd et al., 2010. RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. Mol Cell Biol. 30(10):2460-72

Islam MN, Paquet N et al., 2012. A variant of the BC type 2 susceptibility protein (BRC) repeat is essential for the RECQL5 helicase to interact with RAD51 recombinase for genome stabilization. J Biol Chem. 287 (28): 23808-18

Jalkh N, Chouery E et al., 2017. Next-generation sequencing in familial breast cancer patients from Lebanon. BMC Med Genomics. 10(1): 8

Jemal A, Siegel R et al., 2010. Cancer Statistics, 2010. CA Cancer J Clin. 62(5): 283-98

Johannsdottir HK, Jonsson G et al., 2006. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int J Cancer. 119(5):1052-60

Jönsson G, Naylor TL et al., 2005. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res. 65(17):7612-21

Kakinuma N, Zhuy Y et al., 2009. Kank proteins: structure, functions and diseases. Cell Mol Life Sci. 66(16):2651-9

Kamps R, Brandao RD et al., 2017. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. Int J Mol Sci. 18(2)

Kanagaraj R, Huehn D et al., 2010. RECQL5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. Nucleic Acids Res. 38 (22): 8131-40

Kassube A, Jinek M et al., 2013. Structural mimicry in transcription regulation of human RNA polymerase II by DNA helicase RECQL5. Nat Struct Mol Biol. 20(7):892-9

Kara M, Kaplan M et al., 2016. MTUS1 tumor suppressor and its miRNA regulators in fibroadenoma and breast cancer. Gene. 587(2):173-7

Kataoka K, Nagata Y et al., 2015. Integrated molecular analysis of adult T cell leukemia/lymphoma. Nat Genet. 47(11): 1304-15

Kaufman DJ, Beaty TH et al., 2003. Segregation analysis of 231 Ashkenazi Jewish families for evidence of additional breast cancer susceptibility genes. Cancer Epidemiol Biomarkers Prev. 12(10): 1045-52

Kent T, Chandramouly G et al., 2015. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. Nat Struct Mol Biol. 22(3):230-7

Kiiski JI, Pelttari LM et al., 2014. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. PNAS 111 (42): 15172-7

Kim TM, Jung SH et al., 2014. The mutational burdens and evolutionary ages of early gastric cancers are comparable to those of advanced gastric cancers. J Pathol. 234 (3): 365-74

Kim TM, Son MY et al., 2014. RECQL5 and BLM exhibit divergent functions in cells defective for the Fanconi anemia pathway. Nucleic Acids Res. 43 (2): 893-903

Kim YC, Soliman AS et al., 2017. Unique Features of Germline Variation in Five Egyptian Familial Breast Cancer Families Revealed by Exome Sequencing. PLoS One. 12(1): e0167581

Konitsiotis AD, Jovanovic B et al., 2015. Topological analysis of Hedgehog acyltransferase, a multipalmitoylated transmembrane protein. J Biol Chem. 290(6):3293-307

Kremmidiotis G, Baker E et al., 1998. Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity. Genomics. 49(3):467-71

Ku CS, Naidoo N et al., 2011. Revisiting Mendelian disorders through exome sequencing. Hum Genet. 129(4): 351-70

Kuligina ES, Sokolenko AP et al., 2013. Value of bilateral breast cancer for identification of rare recessive at-risk alleles: evidence for the role of homozygous *GEN1* c.2515_2519delAAGTT mutation. Familial Cancer. 12(1):129-32

Kumar P, Aggarwal R. 2015. An overview of triple-negative breast cancer. Arch Gynecol Obstet. 293(2):247-69

Kurian AW, Hare EE et al., 2014. Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. J Clin Oncol. 32 (19): 2001-9

Lee IH, Sohn M et al., 2014. Ahnak functions as a tumor suppressor via modulation of TGFβ/Smad signaling pathway.Oncogene. 33(38):4675-84

Lemercier C, Duncliffe K et al., 2000. Involvement of Retinoblastoma Protein and HBP1 in Histone $H1^0$ Gene Expression. Mol Cell Biol. 20(18): 6627-37

Levy-Lahad E, Catane R et al., 1997. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. Am J Hum Genet. 60(5): 1059–1067

Li H, Bian C et al., 2011. miR-17-5p promotes human breast cancer cell migration and invasion through suppression of HBP1. Breast Cancer Res Treat. 126(3):565-75

Lin PH, Kuo WH et al., 2016. Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. Oncotarget. 7(7):8310-20

Lindor NM, Furuichi Y et al., 2000. Rothmund-Thomson syndrome due to RECQ4 helicase mutations: report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome. Am J Med Genet. 90(3): 223-8

Lippert H. 2005. Anatomy with clinic orientation. 4a ed. Marbán Editors, Spain. 777 pp

Liu CG, Maercker C et al., 1996. Human plectin: organization of the gene, sequence analysis, and chromosome localization (8q24). PNAS. 93(9):4278-83

Lord CJ, Ashworth A. 2017. PARP inhibitors: Synthetic lethality in the clinic. Science. 355(6330):1152-58

Loveday C, Turnbull C et al., 2011. Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet. 43(9):879-82

Ly D, Forman D et al., 2013. An international comparison of male and female breast cancer incidence rates. Int J Cancer. 132(8): 1918-26

Lynch H, Wen H et al., 2013. Can unknown predisposition in familial breast cancer be family-specific? Breast J. 19 (5): 520-8

Määttä K, Rantapero T et al., 2017. Whole-exome sequencing of Finnish hereditary breast cancer families. Eur J Human Genet. 25(1): 85-93

Macias H, Hinck L. 2012. Mammary gland development. Wiley Interdiscip Rev Dev Biol. 1(4): 533-57

Malhotra GK, Zhao X et al., 2010. Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther. 10(10):955-60

Manolio TA. 2010. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 363 (2): 166-76

Martin AM, Weber BL. 2000. Genetic and Hormonal Risk Factors in Breast Cancer. J Nat Cancer Institute. 92 (14): 1126-35

Marshesi S, Montani F et al., 2014. DEPDC1B coordinates de-adhesion events and cell-cycle progression at mitosis. Dev Cell. 31(4):420-33

Matamala-Zamarro N, 2015. Deregulated microRNAs in breast cancer and their potential role as diagnostic and prognostic biomarkers. PhD Thesis. Madrid, Spain. 160 pp.

Matevossian A, Resh MD. 2015. Hedgehog Acyltransferase as a target in estrogen receptor positive, HER2 amplified, and tamoxifen resistant breast cancer cells. Mol Cell. 14:72

Mavaddat N, Antoniou AC et al., 2010. Genetic susceptibility to breast cancer. Mol Oncol. 4(3):174-91

Mavaddat N, Peock S et al., 2013. Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst. 105(11):812-22

Mavaddat N, Pharoah PD et al., 2015. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 107(5). pii: djv036

McPherson K, Steel CM et al., 2000. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ. 321(7261):624-8

Mehta PA, Tolar J. 2002. Fanconi Anemia in Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Ledbetter N, Mefford HC, Smith RJH, Stephens K, editors. GeneReviews. Washington, US.

Meijers-Heijboer H, van den Ouweland A et al., 2002. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 31 (1): 55-9

Meindl A, Hellebrand H et al., 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 42 (5): 410-4

Melchor L, Benitez J. 2013. The complex genetic landscape of familial breast cancer. Hum Genet. 132 (8): 845-63

Metzker ML. 2010. Sequencing technologies – the next generation. Nat Rev Genet. 11(1): 31-46

Michailidou K, Hall P et al., 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 45 (4): 353-61

Miki Y, Swensen J et al., 1994. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science. 266 (5182): 66-71

Milne RL, Osorio A et al., 2008. The average cumulative risks of breast and ovarian cancer for carriers of mutations in BRCA1 and BRCA2 attending genetic counseling units in Spain. Clin Cancer Res. 14(9):2861-9

Moldovan GL, D'Andrea AD. 2009. How the Fanconi Anemia pathway guards the genome. Annu Rev Genet. 43: 223-49

Moukayed M, Grant WB. 2017. The roles of UVB and vitamin D in reducing risk of cancer incidence and mortality: A review of the epidemiology, clinical trials, and mechanisms. Rev Endocr Metab Disord. 18(2):167-182

Muranen TA, Mavaddat N et al., 2016. Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. Breast Cancer Res Treat. 158(3):463-9

National Cancer Institute, 2016. Breast Cancer Treatment-Patient Version. Image available at www.cancer.gov/types/breast/patient/breast-treatment-pdq, last accessed March 2017

National Comprehensive Cancer Network (NCCN). 2006. Treatment guidelines for patients. American Cancer Society.

Netter FH. 2007. Atlas of human anatomy. 4th ed. Elsevier Masson, Spain. 631 pp

Newman JA, Aitkenhead H et al., 2017. Insights into the RecQ helicase mechanism revealed by the structure of the helicase domain of human RECQL5. Nucleic Acid Res. 45(7):4231-43

Nielsen R, Paul JS et al., 2011. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 12(6):443-51

Nielsen FC, van Overseem Hansen T et al., 2016. Hereditary breast and ovarian cancer: new genes in confined pathways. Nat Rev Cancer. 16 (9): 599-612

Niwa T, Saito H et al., 2009. BRCA2 interacts with the cytoskeletal linker protein plectin to form a complex controlling centrosome localization. Cancer Sci. 100(11):2115-25

Noh JM, Kim J et al., 2015. Exome sequencing in a breast cancer family without BRCA mutation. Radiat Oncol. 33 (2): 149-54

Ofran Y, Rost B. 2007. ISIS: Interaction sites identified from sequence. Bioinformatics. 23(2): e13-6

Ott J, Wang J et al., 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 16(5): 275-84

Ozcan O, Kara M et al., 2016. MTUS1 and its targeting miRNAs in colorectal carcinoma: significant associations. Tumour Biol. 37(5):6637-45

Paine IS, Lewis MT. 2017. The Terminal End Bud: the Little Engine that Could. J Mammary Gland Biol Neoplasia. 22(2):93-108

Paliwal S, Kanagaraj R et al., 2014. Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing. Nucleic Acids Res. 41(4): 2380-90

Park DJ, Odefrey FA et al., 2011. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. Breast Cancer Res Treat. 130(3): 1043-9

Park DJ, Lesueur F et al., 2012. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 90(4): 734-9

Park DJ, Tao K et al., 2014. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. Cancer Discov. 4(7): 804-15

Parkes EE, Kennedy RD. 2016. Clinical Application of Poly(ADP-Ribose) Polymerase Inhibitors in High-Grade Serous Ovarian Cancer. Oncologist. 21(5):586-93

Paulson KE, Rieger-Christ K et al., 2007. Alterations of the HBP1 transcriptional repressor are associated with invasive breast cancer. Cancer Res. 67(13):6136-45

Pearson TA, Manolio TA 2008. How to interpret a genome-wide association study. JAMA. 299(11):1335-44

Pellatt AJ, Wolff RK et al., 2013. Telomere length, telomere-related genes and BC risk: the BC health disparities study. Genes Chromosomes Cancer. 52(7): 595-609

Pennington KP, Walsh T et al., 2014. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. Clin Cancer Res 20 (3):764-75

Perou CM, Sorlie T et al., 2000. Molecular portraits of human breast tumours. Nature. 406 (6797):747-52

Petersen BS, Fredrich B et al., 2017. Opportunities and challenges of whole-genome and -exome sequencing. BMC Genet. 18 (1): 14

Pharoah PD, Antoniou A et al., 2002. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet. 31(1): 33-6

Pharoah PD, Guilford P et al., 2001. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 121(6):1348-53

Pinto R, De Summa S et al., 2014. MicroRNA expression profiling in male and female familial breast cancer. Br J Cancer. 111(12):2361-8

Polyak K. 2007. Breast cancer: originis and evolution. J Clin Invest. 17(11): 3155-63

Popuri V, Tadokoro T et al., 2013. Human RECQL5: guarding the crossroads of DNA replication and transcription and providing backup capability. Crit Rev Biochem Mol Biol. 48(3): 289-99

Poulsen ML, Bisgaard ML. 2008. MUTYH Associated Polyposis (MAP). Curr Genomics. 9(6):420-35

Prat A, Parker JS et al., 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 12(5):R68

Qi Y, Zhou X. 2014. Haplotype analysis of RECQL5 gene and laryngeal cancer. Tumour Biol. 35(3): 2669-73

Rabbani B, Mahdieh N et al., 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet. 57(10): 621-32

Rabbani B, Tekin M et al., 2014. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 59 (1): 5-15

Rahman N, Seal S et al., 2007. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 39 (2): 165-7

Rahman N. 2014. Realizing the promise of cancer predisposition genes. Nature. 505 (7483): 302-8

Rainville IR, Rana HQ. 2014. Next-Generation sequencing for inherited breast cancer risk: counseling through the complexity. Curr Oncol Rep 16 (3):371

Ramus SJ, Song H et al., 2015. Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. J Natl Cancer Inst. 107(11)

Ray S, Pollard JW. 2012. KLF15 negatively regulates estrogen-induced epithelial cell proliferation by inhibition of DNA replication licensing. PNAS. 109(21):E1334-43

Renwick A, Thompson D et al., 2006. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 38 (8):873-75

Richards S, Aziz N et al., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Met. 17(5): 405-24

Rizzolo P, Silvestri V et al., 2013. Male breast cancer: genetics, epigenetics and ethical aspects. Ann Oncol. 24 Suppl 8:viii75-82

Robert M, Frenel JS et al., 2017. Olaparib for the treatment of breast cancer. Expert Opin Investig Drugs. 26(6):751-9

Roberts DG, Morrison TB et al., 2009. A nanoliter fluidic platform for large-scale single nucleotide polymorphism genotyping. Biotechniques.46 (3Suppl):ix-xiii

Rodrigues-Ferreira S, Di Tommaso A et al., 2009. 8p22 MTUS1 gene product ATIP3 is a novel anti-mitotic protein underexpressed in invasive breast carcinoma of poor prognosis. PLoS One. 4(10):e7239

Rosa-Rosa JM, Pita G et al., 2009. Genome-wide linkage scan reveals three putative breast-cancer-susceptibility loci. Am J Hum Genet. 84(2): 115-22

Rosa-Rosa JM, Gracia-Aznárez FJ et al., 2010. Deep sequencing of target linkage assay-identified regions in familial breast cancer: methods, analysis pipeline and troubleshooting. PLoS One. 5(4): e9976

Ross MH, Pawlina W. 2013. Histology: text and color atlas with molecular and cellular biology. Panamerican Editors. 991 pp

Roy BC, Aoyagi T et al., 2005. Pathological characterization of Kank in renal cell carcinoma. Exp Mol Pathol. 78(1):41-8

Rudolph A, Chang-Claude J et al., 2016. Gene–environment interaction and risk of breast cancer. Br J Cancer. 114 (2): 125-33

Sampson EM, Haque ZK et al., 2001. Negative regulation of the Wnt-β catetin pathway by the transcriptional repressor HBP1. EMBO J 20(16):4500-11

Saponaro M, Kantidakis T et al., 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell. 157(5):1037-49.

Sarkar S, Roy BC et al., 2002. A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. 277(39):36585-91

Sas-Korczynska B, Adamczyk A et al., 2015. Androgen receptor in male breast cancer. Pol J Pathol. 66(4):347-52

Schlessinger A, Rost B. 2005. Protein flexibility and rigidity predicted from sequence. Proteins: Structure, Function and Genetics. 61 (1): 115-26

Schrader KA, Masciari S et al., 2008. Hereditary diffuse gastric cancer: association with lobular breast cancer. Fam Cancer. 7 (1): 73-82

Schmidt M, Voelker HU et al., 2010. Glycolytic phenotype in breast cancer: activation of Akt, up-regulation of GLUT1, TKTL1 and down-regulation of M2PK. J Cancer Res Clin Oncol. 136(2):219-25

Schmidt T, Karsunky H et al., 2000. A novel protein (Fbf-1) that binds to CD95/APO-1/FAS and shows sequence similarity to trichohyalin and plectinc. Biochim Biophys Acta. 1493(1-2):249-54

Schnitt SJ. 2010. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. Modern pathology: an official journal of the United States and Canadian Academy of Pathology. 23(2): S60-4

Schwendener S, Raynard S et al., 2010. Physical interaction of RECQ5 helicase with RAD51 facilitates its anti-recombinase activity. J Biol Chem. 285(21):15739-45

Seal S, Thompson D et al., 2006. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet. 38(11):1239-41

Shamanna RA, Lu H et al., 2016. Camptothecin targets WRN protein: mechanism and relevance in clinical breast cancer. 7(12):13269-84

Sharp TV, Munoz F et al., 2004. LIM domains-containing protein 1 (LIMD1), a tumor suppressor encoded at chromosome 3p21.3, binds pRB and represses E2F-driven transcription. PNAS. 101(47):16531-6

Shtivelman E, Cohen FE et al., 1992. A human gene (AHNAK) encoding an unusually large protein with a 1.2-microns polyionic rod structure. PNAS. 89(12):5472-6

Siegel RL, Miller KD et al., 2017. Cancer statistics, 2017. CA Cancer J Clin. 67(1): 7-30

Silva TA, Smuczek B et al., 2016. AHNAK enables mammary carcinoma cells to produce extracellular vesicles that increase neighboring fibroblast cell motility. Oncotarget. 7(31):49998-50016

Simon R, Roychowdhury S. 2013. Implementing personalized cancer genomics in clinical trials. Nat Rev Drug Discov. 12(5):358-69

Singletary SE, 2003. Rating the risk factors for breast cancer. Ann Surg. 237(4): 474- 82

Smith A, Moran A et al., 2007. Phenocopies in BRCA1 and BRCA2 families: evidence for modifier genes and implications for screening. J Med Genet. 44(1): 10-15

Snape K, Ruark E et al., 2012. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res Treat. 134(1): 429-33

Song H, Dicks E et al., 2015. Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. J Clin Oncol. 33(26):2901-7

Sorlie T, Perou CM et al., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. PNAS. 98(19):10869-74

Southey MC, Goldgar DE et al., 2016. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet. 53 (12): 800-11

Sparano JA, Lee MC. 2016. Breast Cancer Staging. MedScape. http://emedicine.medscape.com/article/2007112-overview. Last accessed April 2017

Speirs V, Shaaban AM. 2009. The rising incidence of male breast cancer. Breast Cancer Res Treat. 115(2):429-30

Steel M, Thompson A et al., 1991. Genetic aspects of breast cancer. Br Med Bull. 47(2):504-18

Sternlicht MD. 2006. Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res. 8(1): 201

Stiff T, Shtievelman E et al., 2004. AHNAK interacts with the DNA ligase IV-XRCC4 complex and stimulates DNA ligase IV-mediated double-stranded ligation. DNA Repair (Amst). 3(3):245-56

Stratton MR, Rahman N. 2008. The emerging landscape of breast cancer susceptibility. Nat Genet. 40(1): 17-22

Sujimoto M, Inoko A et al., 2008. The keratin-binding protein Albatross regulates polarization of epithelial cells.. J Cell Biol. 183(1):19-28

Sun J, Wang Y et al., 2015. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet. 11(5): e1005228

Suzuki JI, Roy BC et al., 2017. Depletion of tumor suppressor Kank1 induces centrosomal amplification via hyperactivation of RhoA. Exp Cell Res. 353(2):79-87

Tabor HK, Risch NJ et al., 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 3(5):391-7

Tan MH, Mester JL et al., 2012. Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res. 18(2): 400-7

Tang J, Cho NW et al., 2013. Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. Nat Struct Mol Biol. 20(3):317-25

Taylor AM, Lam Z et al., 2015. Ataxia telangiectasia: more variation at clinical and cellular levels. Clin Genet 87 (3):199-208

Telatar M, Wang Z et al., 1996. Ataxia-telangiectasia: mutations in ATM cDNA detected by protein-truncation screening. Am J Hum Genet 59 (1):40-4

Thompson ER, Doyle MA et al., 2012. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. Plos Genet. 8 (9):e1002894

Thompson ER, Rowley SM et al., 2013. Analysis of RAD51D in ovarian cancer patients and families with a history of ovarian or breast cancer. PLoS One. 8(1):e54772

Torres LA, Siegel RL, et al., 2016. Global cancer incidence and mortality rates and trends. An update. Cancer Epidiomol Biomarkers Prev. 25(1): 16-27

Tournier I, Paillerets BB et al., 2004. Significant contribution of germline BRCA2 rearrangements in male breast cancer families. Cancer Res. 64 (22):8143-7

Townsend CM, Beauchamp RD et al., 2013. Sabiston. Surgery Manifest. 19th. Elsevier Masson, Editors. Spain. 2144 pp

Tung N, Domchek SM et al., 2016. Counselling framework for moderate-penetrance cancer-susceptibility mutations. Nat Rev Clin Oncol. 13(9):581-8

Turnbull C, Rahman N. 2008. Genetic predisposition to breast cancer: past, present and future. Annu Rev Genomics Hum Genet. 9: 321-345

Turnbull C, Hines S et al., 2010. Mutation and association analysis of GEN1 in breast cancer susceptibility. Breast Cancer Res Treat. 124(1):283-8

Van Allen EM, Wagle N et al., 2014. The Genetic Landscape of Clinical Resistance to RAF Inhibition in Metastatic Melanoma. Cancer Discov. 4 (1): 94-109

Van der Groep P, van der Wall E et al., 2011. Pathology of hereditary breast cancer. Cell Oncol. 34 (2):71-88

van Lier MG, Wagner A et al., 2010. High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol. 105(6):1258-64

Verardo LL, Silva FF et al., 2016. Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. Genet Sel Evol. 48:9

Verhagen MM, Last JI et al., 2012. Presence of ATM protein and residual kinase activity correlates with the phenotype in ataxia-telangiectasia: a genotype-phenotype study. Hum Mutat 33 (3):561-71

Wang X, Lu X et al., 2011. RECQL5 is an important determinant for camptothecin tolerance in human colorectal cancer cells.. Biosci Rep. 31(5):363-9

Wang K, Ren Y et al., 2017. miR-4262 Promotes Proliferation and Invasion of Human Breast Cancer Cells Through Directly Targeting KLF6 and KLF15. Oncol Res. 25(2):277-283

Wen H, Kim YC et al., 2014. Family-specific, novel, deleterious germline variants provide a rich resource to identify genetic predispositions for BRCAX familial breast cancer. BMC Cancer. 14:470

Wiche G, Hermann H et al., 1982. Plectin: a high-molecular-weight cytoskeletal polypeptide component that copurifies with intermediate filaments of the vimentin type. Cold Spring Harb Symp Quant Biol. 1:475-82

Wong E, Rebelo J. 2012. McMaster pathophysiology Review. Information available in: http://www.pathophys.org/breast-cancer/breastcancer-copy/, last accessed April 2017

Wooster R, Mangion J et al 1992. A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. Nat Genet. 2(2):132-4

Wooster R, Bignell G et al., 1995. Identification of the breast cancer susceptibility gene *BRCA2*. Nature. 378 (6559):789-92

Xuan J, Yu Y et al., 2013. Next-generation sequencing in the clinic: Promises and challenges. Cancer Lett. 340(2):284-95

Yachdav G, Kloppmann E et al., 2014. PredictProtein-an open resource for online prediction of protein structural and functional features. Nucleic Acid Res. 42: W337–43

Yoda T, McNamara KM et al., 2015. KLF15 in breast cancer: a novel tumor suppressor? Cell Oncol (Dordr). 38(3):227-35

Zhang X, Kim J et al., 2006. The HBP1 transcriptional repressor participates in RAS-induced premature senescence. Mol Cell Biol. 26(22):8252-66

Zhang Y, Subbaiah VK et al., 2016. TIP60 inhibits metastasis by ablating DNMT1-SNAIL2-driven epithelial-mesenchymal transition program. J Mol Cell Bio . pii: mjw038

Zhao T, Ding X et al., 2015. MTUS1/ATIP3a down-regulation is associated with enhanced migration, invasion and poor prognosis in salivary adenoid cystic carcinoma. BMC Cancer. 15:203

Zhi LQ, Ma W et al., 2014. Association of RECQL5 gene polymorphisms and osteosarcoma in a Chinese Han population. Tumour Biol. 35(4): 3255-59

Zhu M, Zhao S. 2007. Candidate gene identification approach: progress and challenges. Int J Biol Sci. 3(7): 420-7

10. Publications and Conference Posters

Publications

Tavera-Tapia A, Pérez-Carbonero L, Macías JA, Ceballos MI, Roncador G, de la Hoya M, Barroso A, Felipe-Ponce V, Serrano-Blanch R, Hinojo C, Miramar-Gallart MD, Urioste M, Caldés T, Santillán-Garzón S, Benitez J, Osorio A. (2017). Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the *ATM* gene. Breast Cancer Res Treat. 161(3): 597-604

Tavera-Tapia A, Calvete O, de la Hoya M, Macías JA, Alonso B, Alonso MR, Pita G, Barroso A, Fernández V, Urioste M, Caldés T, Benítez J, Osorio A. *RECQL5*: another DNA helicase potentially involved in breast cancer susceptibility. Manuscript in preparation.

Posters in Conferences

Search for new high susceptibility genes in hereditary breast cancer. **Tavera A**, de la Hoya M, Macías A, Barroso A, Urioste M, Benitez J, Osorio A. European Congress of Human Genetics. Barcelona, Spain. May 2016.

Looking for novel susceptibility genes in hereditary male breast cancer. **Tavera A,** de la Hoya M, Barroso A, Urioste M, Caldés T, Benitez J, Osorio A. Fifth Familial Cancer Conference. 6 ECTS. Madrid, Spain. May 2016.

Quest of novel high susceptibility hereditary breast cancer genes in families with a recessive pattern of inheritance: WES approach. **Tavera A**, de la Hoya M, Macías A, Barroso A, Urioste M, Benitez J, Osorio A. CNIO PhDay. Madrid, Spain. December 2015.

Looking for novel high susceptibility genes in hereditary breast cancer families with a recessive pattern of inheritance. **Tavera A**, de la Hoya M, Macías A, Barroso A, Urioste M, Benitez J, Osorio A. CNIC PhDay. Madrid, Spain. November 2015.

Search for new high susceptibility genes in hereditary breast cancer families with a recessive pattern of inheritance. **Tavera A**, de la Hoya M, Macías A, Barroso A, Urioste M, Benitez J, Osorio A. Nordic Next Generation Sequencing Congress. Lund, Sweden. October 2015.